PA NT COOPERATION TREAT

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

To:

Commissioner
US Department of Commerce
United States Patent and Trademark
Office, PCT
2011 South Clark Place Room
CP2/5C24
Arlington, VA 22202
ETATS-UNIS D'AMERIQUE

in its capacity as elected Office

Date of mailing (day/month/year) 12 July 2001 (12.07.01)	
International application No. PCT/IL00/00736	**Applicant's or agent's file reference** 129275.4 DAB
International filing date (day/month/year) 10 November 2000 (10.11.00)	**Priority date** (day/month/year) 10 November 1999 (10.11.99)
Applicant TOPOROIK, Amir et al	

1. The designated Office is hereby notified of its election made:

   [X] in the demand filed with the International Preliminary Examining Authority on:

   21 May 2001 (21.05.01)

   [ ] in a notice effecting later election filed with the International Bureau on:

2. The election [X] was

   [ ] was not

   made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO **34, chemin des Colombettes** **1211 Geneva 20, Switzerland**	Authorized officer Odile ALIU
Facsimile No.: (41-22) 740.14.35	Telephone No.: (41-22) 338.83.38

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 May 2001 (17.05.2001)

PCT

(10) International Publication Number
WO 01/34796 A1

(51) International Patent Classification[7]: C12N 15/12, C07K 14/475, 16/22, C12Q 1/68, G01N 33/68, 33/53, A61K 38/17

(21) International Application Number: PCT/IL00/00736

(22) International Filing Date: 10 November 2000 (10.11.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

132846	10 November 1999 (10.11.1999)	IL
133767	28 December 1999 (28.12.1999)	IL

(71) Applicant *(for all designated States except US)*: **COMPUGEN LTD.** [IL/IL]; Pinchas Rozen Street 72, 69512 Tel Aviv (IL).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TOPOROIK, Amir** [IL/IL]; Kazanelson Street 20/2, 58016 Azur (IL). **BITON, Sharon** [IL/IL]; Ilanot Street 71a, 46565 Herzliya (IL). **SAVITZKY, Kinneret** [IL/IL]; Metodela Street 44, 69548 Tel Aviv (IL). **BERNSTEIN, Jeanne** [IL/IL]; Harimon Street 23, 40300 Kfar Yona (IL).

(74) Agent: **REINHOLD COHN AND PARTNERS**; P.O.B. 4060, 61040 Tel Aviv (IL).

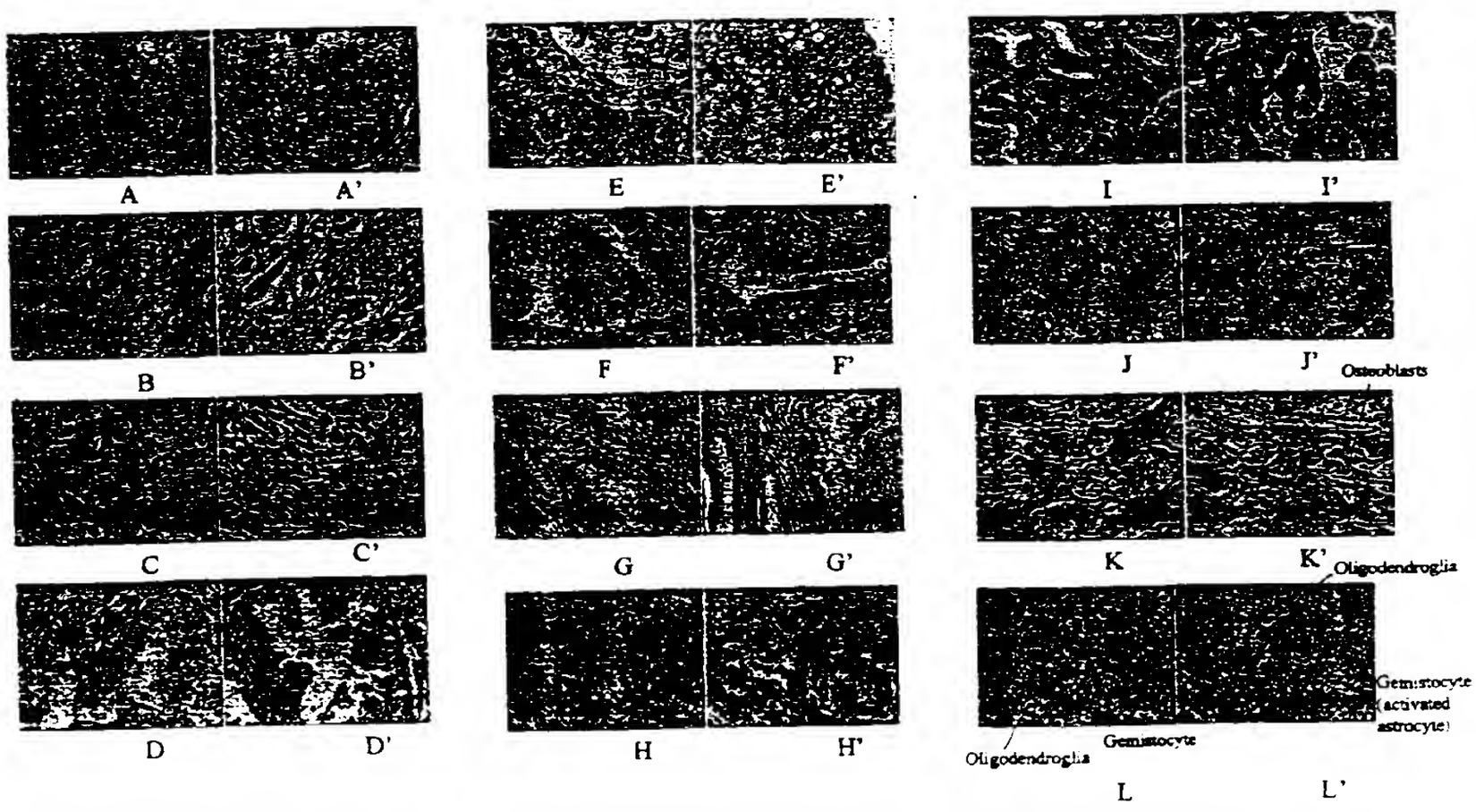
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*

[Continued on next page]

(54) Title: CHORDIN-LIKE HOMOLOGS



(57) Abstract: The present invention concerns several splice variants of a chordin line homologs (CLH) and depicts their nucleic acid and amino acid sequences, vectors and host cells containing said nucleic acid sequences and antibodies reactive with the amino acid sequences. The invention also concerns pharmaceutical compositions for the treatment of a plurality of diseases, comprising nucleic acid sequences, amino acid sequences, expression vectors, antibodies. The invention also concerns methods for detecting the above nucleic acid, amino acid sequences or antibody in a biological sample.

WO 01/34796 A1

-- *Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

CHORDIN-LIKE HOMOLOGS

FIELD OF THE INVENTION

The present invention concerns novel nucleic acid sequences, vectors and host cells containing them, amino acid sequences encoded by said sequences, and antibodies reactive with said amino acid sequences, as well as pharmaceutical compositions comprising any of the above. The present invention further concerns methods for screening for candidate activator or deactivators utilizing said amino acid sequences.

BACKGROUND OF THE INVENTION

The TGFß superfamily is composed of a range of functional and structural factor subclasses with predominantly growth-inhibitory cellular actions and developmental regulatory effects on organogenesis, pattern formation, modulation of extracellular matrix and terminal differentiation. The subclasses include the TGFß, activins, glial-derived factors (GDFs), Mullerian inhibiting substances, glial-derived neurotrophic factor (GDNF), cartilage-derived morphogenetic proteins (CDMPs) and the rapidly expanding factor subclass of bone morphogenic proteins (BMPs). BMPs participate in a broad spectrum of cellular inducing events involving all three germ layers during metazoan development. There are now known to be 7 members of this family (BMPs 1-7); all except BMP1 are members of the TGF-α family. BMP1 has been classed as a novel regulatory protein. The term 'bone morphogenetic' may, however, prove to be a misnomer, since the messenger RNA for the BMPs are expressed in a wide variety of tissues, suggesting limited tissue specificity of function.

Chordin is an abundant glycoprotein with molecular mass of 120Kda. It contains internal cystein rich repeats called Von Willbrand domain and N-glycosylation sites.

Chordin is a key developmental protein that dorsalizes early vertebrate embryonic tissues by binding to ventralizing TGF-beta -like bone morphogenetic proteins (BMP) and sequestering them in latent complexes. Chordin binds to ventral BMP-2 and BMP-4 signal in the extracellular space, blocking the interaction of BMPs with their receptors. Chordin mimics the action of the Spemann organizer and can induce the formation of neural tissue from ectoderm and dorsalization of the ventral mesoderm to form muscle.

GLOSSARY

In the following description and claims use will be made, at times, with a variety of terms, and the meaning of such terms as they should be construed in accordance with the invention is as follows:

"Chordin like homolog (CLH) nucleic acid sequence" – the sequence shown in any one of SEQ ID NO: 1 to 11, sequences having at least 70% identity to said sequence and fragments of the above sequences being 20 b.p. long. Those sequences are sequences coding for a novel homolog of the known Chordin protein, as well as for variants of the novel homolog produced by alternative splicing.

The sequence shown in SEQ ID NO: 1 is a homolog to the known chordins within the VWFC domain, named after the von-Willebrand factor (VWF) type C repeat, which is found 2-4 times in these multi-domain proteins. The VWF domain has a length of about 70 amino acids covering 10 well conserved cysteines. The presence of this region in complex-forming proteins leads to the assumption that the VWFC domain might be involved in forming larger protein complexes. The homolog is a part of a longer sequence termed hereinafter *"full sequence"*. The full sequence has naturally occurring splice variants which are also termed CLH. The first variant (SEQ ID NO: 2) has 3 out of VWFC domains of the known

chordin. The protein coded therefrom contains a predicted signal peptide. The second variant (SEQ ID NO: 3) and third variant (SEQ ID NO: 4) contain 3 out of 4 VWFC domains of the known chordin, but is not predicted to contain the signal sequence. Sequences SEQID 5 to SEQ ID NO.10 are also splice variants of the full sequence. SEQ ID NO. 5 contains 3 out of 4 VWFC domains of known chordins (domain #2,3,4). The VWFC domain is named after the von Willebrand factor (VWF) type C repeat which is found twice in this multidomain protein. It has a length of about 70 amino acids covering 10 well conserved cysteines. The protein of chordin-like variant 1 at SEQ ID 5 (depicted in SEQ ID 16) contains predicted signal peptide. SEQ ID NO. 6 contains 3 out of the 4 VWFC domains of known chordin and the protein encoded thereby (SEQ ID 17), contains a predicted signal peptide. SEQ ID NO. 7 contains 2 out of 4 VWFC domains of known chordin. SEQ ID NO. 8 has out of the 4 VWF Factor Type C domains. SEQ ID NO. 9 has 2 VWF and SEQ ID NO. 10 has 2 VWF domains. SEQ ID No. 11 is a mouse ortolog of the CLH of the inventor.

However, the term CLH does not necessarily signify that CLH protein coded by the above sequences (including the variant sequences) has the same or even similar physiological effects as known Chordins, merely that it shows sequence homology with the known Chordin.

"Variant" – a sequence produced by alternative splicing of full sequence homolog. These sequences are not merely truncated forms of the full sequence, or modifications of the full sequence, but rather naturally occurring sequences resulting from various alternative splicings.

"Chordin like homology product (CLH product) – also referred at times as the "CLH protein" or ***"CLH polypeptide"*** – is an amino acid coded by any one of SEQ ID NOS: 1 to 11. The amino acid sequence may be a peptide, a protein, as well as peptides or proteins having *chemically modified* amino acids (see below) such as a glycopeptide or glycoprotein. An example of an CLH product is shown in any one of SEQ ID NOS: 12 to 22. The term also includes analogues of said sequences in which one or more amino acids has been added, deleted, *substituted*

(see below) or *chemically modified* (see below) as well as fragments of this sequence having at least 10 amino acids.

"Nucleic acid sequence" – a sequence composed of DNA nucleotides, RNA nucleotides or a combination of both types and may includes natural nucleotides, chemically modified nucleotides and synthetic nucleotides.

"Amino acid sequence" – a sequence composed of any one of the 20 naturally appearing amino acids, amino acids which have been *chemically modified* (see below), or composed of synthetic amino acids.

"Fragment of CLH product" - a polypeptide which has an amino acid sequence which is the same as part of but not all of the amino acid sequence of the CLH product.

"Fragments of CLH nucleic acid sequence" a continuous portion, preferably of about 20 nucleic acid sequences of the CLH nucleic acid sequence.

"Conservative substitution" - refers to the substitution of an amino acid in one class by an amino acid of the same class, where a class is defined by common physicochemical amino acid side chain properties and high substitution frequencies in homologous proteins found in nature, as determined, for example, by a standard Dayhoff frequency exchange matrix or BLOSUM matrix. [Six general classes of amino acid side chains have been categorized and include: Class I (Cys); Class II (Ser, Thr, Pro, Ala, Gly); Class III (Asn, Asp, Gln, Glu); Class IV (His, Arg, Lys); Class V (Ile, Leu, Val, Met); and Class VI (Phe, Tyr, Trp). For example, substitution of an Asp for another class III residue such as Asn, Gln, or Glu, is a conservative substitution.

"Non-conservative substitution" - refers to the substitution of an amino acid in one class with an amino acid from another class; for example, substitution of an Ala, a class II residue, with a class III residue such as Asp, Asn, Glu, or Gln.

"Chemically modified" - when referring to the product of the invention, means a product (protein) where at least one of its amino acid resides is modified either by natural processes, such as processing or other post-translational modifications, or by chemical modification techniques which are well known in the art. Among the numerous known modifications typical, but not exclusive examples include: acetylation, acylation, amidation, ADP-ribosylation, glycosylation, GPI anchor formation, covalent attachment of a lipid or lipid derivative, methylation, myristlyation, pegylation, prenylation, phosphorylation, ubiqutination, or any similar process.

"Biologically active" - refers to the CLH product which have, regulatory or biochemical functions on the same target sites which naturally occurring CLH influence, for example can bind to the same receptor as the chordin (or to another receptor).

"Immunologically active" defines the capability of a natural, recombinant or synthetic CLH product, or any fragment thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies. Thus, for example, a biologically active fragment of CLH product denotes a fragment which retains some or all of the immunological properties of the CLH product, e.g can bind specific anti-CLH product antibodies or which can elicit an immune response which will generate such antibodies or cause proliferation of specific immune cells which produce CLH.

"Optimal alignment" - is defined as an alignment giving the highest percent identity score. Such alignment can be performed using a variety of commercially

available sequence analysis programs, such as the local alignment program LALIGN using a ktup of 1, default parameters and the default PAM. A preferred alignment is the one performed using the CLUSTAL-W program from MacVector (TM), operated with an open gap penalty of 10.0, an extended gap penalty of 0.1, and a BLOSUM similarity matrix. If a gap needs to be inserted into a first sequence to optimally align it with a second sequence, the percent identity is calculated using only the residues that are paired with a corresponding amino acid residue (i.e., the calculation does not consider residues in the second sequences that are in the *"gap"* of the first sequence).

"Having at least X% identity" - with respect to two amino acid or nucleic acid sequence sequences, refers to the percentage of residues that are identical in the two sequences when the sequences are optimally aligned. Thus, 70% amino acid sequence identity means that 70% of the amino acids in two or more optimally aligned polypeptide sequences are identical.

"Isolated nucleic acid molecule having an CLH nucleic acid sequence" - is a nucleic acid molecule that includes the coding CLH nucleic acid sequence. Said isolated nucleic acid molecule may include the CLH nucleic acid sequence as an independent insert; may include the CLH nucleic acid sequence fused to an additional coding sequences, encoding together a fusion protein in which the CLH coding sequence is the dominant coding sequence (for example, the additional coding sequence may code for a signal peptide); the CLH nucleic acid sequence may be in combination with non-coding sequences, e.g., introns or control elements, such as promoter and terminator elements or 5' and/or 3' untranslated regions, effective for expression of the coding sequence in a suitable host; or may be a vector in which the CLH protein coding sequence is a heterologous.

"Expression vector" - refers to vectors that have the ability to incorporate and express heterologous DNA fragments in a foreign cell. Many prokaryotic and eukaryotic expression vectors are known and/or commercially available. Selection of appropriate expression vectors is within the knowledge of those having skill in the art.

"Deletion" - is a change in either nucleotide or amino acid sequence in which one or more nucleotides or amino acid residues, respectively, are absent.

"Insertion" or "addition" - is that change in a nucleotide or amino acid sequence which has resulted in the addition of one or more nucleotides or amino acid residues, respectively, as compared to the naturally occurring sequence.

"Substitution" - replacement of one or more nucleotides or amino acids by different nucleotides or amino acids, respectively. As regards amino acid sequences the substitution may be conservative or non- conservative.

"Antibody" – refers to IgG, IgM, IgD, IgA, and IgG antibody. The definition includes polyclonal antibodies or monoclonal antibodies. This term refers to whole antibodies or fragments of the antibodies comprising the antigen-binding domain of the anti-CLH product antibodies, e.g. antibodies without the Fc portion, single chain antibodies, fragments consisting of essentially only the variable, antigen-binding domain of the antibody, etc.

"Activator" - as used herein, refers to a molecule which mimics the effect of the natural CLH product or at times even increases or prolongs the duration of the biological activity of said product, as compared to that induced by the natural product. The mechanism may be by binding to the same receptor of target moieties to which native CLH binds thus mimicking the activity of CLH; by prolonging the lifetime of the CLH, (for example by decrease of the rate of its

degradation) by increasing the activity of the CLH on its target (modulation of expression and amount of BMPs), by increasing the affinity of CLH to moieties which it binds (such as its receptors) etc. Activators may be polypeptides, nucleic acids, carbohydrates, lipids, or derivatives thereof, or any other molecules which can bind to and activate the CLH product.

"Deactivator" or ***("Inhibitor")*** - refers to a molecule which modulates the activity of the CLH product in an opposite manner to that of the activator, by decreasing or shortening the duration of the biological activity of the CLH product. This may be done by blocking the binding of the CLH to its receptor (competitive or non-competitive inhibition), by causing rapid degradation of the CLH, etc. by inhibiting association of the CLH with the effectors which regulate the expression of BMPs, etc. Deactivators may be polypeptides, nucleic acids, carbohydrates, lipids, or derivatives thereof, or any other molecules which bind to and modulate the activity of said product.

"Treating a disease" - refers to administering a therapeutic substance effective to ameliorate symptoms associated with a disease, to lessen the severity or cure the disease, or to prevent the disease from occurring.

"***Detection***" – refers to a method of detection of a disease. This term may refer to detection of a predisposition to a disease.

"Probe" – the CLH nucleic acid sequence, or a sequence complementary therewith, when used to detect presence of other similar sequences in a sample. The detection is carried out by identification of hybridization complexes between the probe and the assayed sequence. The probe may be attached to a solid support or to a detectable label.

SUMMARY OF THE INVENTION

The present invention is based on the surprising finding that there exist in humans (and mice), several novel homologs of the chordin protein having a significant homology to the chordin protein, the homolog is a part of a longer sequence termed *"full sequence"*. The invention is further based on the surprising finding that there exist several splice variants to the full sequence which variants are naturally occurring sequences produced from the novel homolog through alternative splicing. Both the homolog and the variants of the full sequence are collectively termed as *"CLH"*.

The novel CLH (in SEQ ID NO: 1) is a homolog to the known chordins within the VWFC domain, named after the von-Willebrand factor (VWF) type C repeat, which is found 2-4 times in these multi-domain proteins. The VWF domain has a length of about 70 amino acids covering 10 well conserved cysteines. The presence of this region in complex-forming proteins leads to the assumption that the VWFC domain might be involved in forming larger protein complexes. The other variants to the full sequence (for which the homolog is a portion) (SEQ ID NO: 2-10) have 2, 3 or 4 VWF type repeats. SEQ ID NO. 2 and 6 also has a sequence coding for a signal sequence, while SEQ ID NO: 3 and 4 are predicted not to have such a signal sequence.

Thus the present invention provides by its first aspect, a novel isolated nucleic acid molecule comprising or consisting of the sequence of any one of SEQ ID NO: 1 to SEQ ID NO: 11, fragments of said sequence having at least 20 nucleic acids, or a molecule comprising a sequence having at least 70%, preferably 80%, and most preferably 90% or 95% identity to any one of SEQ ID NO:1 to SEQ ID NO: 11.

The present invention further provides a protein or polypeptide comprising or consisting of an amino acid sequence encoded by any of the above nucleic acid sequences, termed herein *"CLH product"*, for example, an amino acid sequence having the sequence as depicted in any one of SEQ ID NO: 12 to 22, fragments of the above amino acid sequence having a length of at least 10 amino acids, as well

as homologs of the amino acid sequences of any one of SEQ ID NO: 12 to 22 in which one or more of the amino acid residues has been substituted (by conservative or non-conservative substitution) added, deleted, or chemically modified.

The present invention further provides nucleic acid molecule comprising or consisting of a sequence which encodes the above amino acid sequences, (including the fragments and analogs of the amino acid sequences). Due to the degenerative nature of the genetic code, a plurality of alternative nucleic acid sequences, beyond SEQ ID NO:1 to SEQ ID NO: 11, can code for the amino acid sequence of the invention. Those alternative nucleic acid sequences which code for the same amino acid sequences codes by the sequences of SEQ ID NO: 1 to SEQ ID NO: 11 are also an aspect of the of the present invention.

The present invention further provides expression vectors and cloning vectors comprising any of the above nucleic acid sequences, as well as host cells transfected by said vectors.

The present invention still further provides pharmaceutical compositions comprising, as an active ingredient, said nucleic acid molecules, said expression vectors, or said protein or polypeptide.

By a second aspect, the present invention provides a nucleic acid molecule comprising or consisting of a non-coding sequence which is complementary to that of any one of SEQ ID NO: 1 to SEQ ID NO: 11, or complementary to a sequence having at least 70%, preferably 80%, most preferably 90% or 95% identity to said sequence or a fragment of said two sequences. The complementary sequence may be a DNA sequence which hybridizes with any one of the sequences of SEQ ID NO: 1 to SEQ ID NO: 11, or hybridizes to a portion of that sequence having a length sufficient to inhibit the transcription of the complementary sequence. The complementary sequence may be a DNA sequence which can be transcribed into an mRNA being an antisense to the mRNA transcribed from any one of SEQ ID NO: 1 to SEQ ID NO: 11 or into an mRNA which is an antisense to a fragment of the mRNA transcribed from any one of SEQ ID NO: 1 to SEQ ID NO: 11 which has a length sufficient to hybridize with the mRNA transcribed from any one of SEQ ID

NO: 1 to SEQ ID NO: 11, so as to inhibit its translation. The complementary sequence may also be the mRNA or the fragment of the mRNA itself. The pharmaceutical compositions of the invention (according to both aspects may be used for the treatment of a plurality of diseases.

In accordance with the present invention, it has been found that the CLH of the invention is located in astrocytes. As astrocytes are known to have a variety of physiological activities in maintaining normal brain physiology, such as in the secretion of active compounds, formation of the blood-brain barrier, metabolism of neurotransmitters and maintenance of the ionic balance of the extracellular space. Pharmaceutical compositions in accordance with the present invention may be used to treat diseases and pathological conditions which can be benefited by a modulation of astrocyte activity, such as the modulation of the cross-talk signals in the CNS during physiological and pathological conditions of the nervous system. Examples of such diseases are neuro-degenerative diseases caused by aging, infectious agents, by toxic substances or due to genetic causes. In addition, the pharmaceutical compositions may be used for the treatment of diseases, and pathological conditions involving abnormal development of the nervous system.

It has been postulated that chordin may be expressed by cells of the osteoblastic lineage to limit BMP actions in the osteoblast. This would be a critical function for a BMP binding protein since excessive BMP-4 has been implicated in pathogenesis of fibrodysplasia ossificans progressiva. Fibrodysplasia Ossificans Progressiva (FOP) is a rare genetic disease in which muscles, tendons, ligaments and other connective tissues may ossify into bone. BMPs can cause induction of noggin and chordin mRNA and protein levels in skeletan cells by trasncriptional mechanisms, and in turn these prevent the effect of BMPs in osteoblast in a negative-type feeback. The induction of these proteins by BMPs appears to be a mechanism to limit the BMP effect in bones. Existing therapies which are being investigated for their effectiveness in preventing heterotopic bone formation include BMP's inhibitors.

Considerable evidence exists supporting a role for TGF in morphogenesis,

in the regulation of endochondral ossification and in bone remodelling. TGF effect the proliferation and differentiation of osteoblastic cells *in vitro* and high levels of messenger RNA are expressed in the growth plate of fetal human long bones.

The CLH of the invention was found, by immunohystochemical methods to be localized in fetal-human bone .

Thus, the pharmaceutical compositions of the present invention may be used for the treatment of diseases and pathological conditions associated with osteoblastists or other diseases of mesanchimal origin. An example of such diseases is Fibrodysplasia Ossificans, as well as other diseases involving abnormal bone or cartilage formation, metabolism and/or destruction.

Furthermore, the CLH variances of the invention were mapped to chromosome 11q14 (genomic clone accession no. APOO 2010; AP001324; ACO118686).

The chromosomal location of the CLH gene is near several disorders of cartilage and bone formation, and thus, the pharmaceutical compositions of the invention may be used for the treatment or alleviation of the following diseases: Osteopetrosis, Autosomal Recessive (congenital disorder characterized also by development of abnormally dense bones).

High Bone Mass (HBM) - High bone mass can result from osteosclerosis (increased density of trabecular - spongy bone) and/or hyperostosis (thickening of cortical - compact bone from deposition of osseous tissue) along subperiosteal and/or endosteal surfaces), occurring focally or throughout the skeleton.

The pharmaceutical compositions of the invention may be used also for the treatment of osteoporosis pseudoglioma syndrome, autosomal recessive osteopetrosis, and isolated increased bone mass (high bone mass without other clinical features). The CLH of the invention may also be used for augmenting bone regeneration after injury, so as to speed up the healing process.

In accordance with the findings of the present invention, CLH of the invention is expressed in the placenta, and is localized in the uterus lining

(endometrium). It is known, that poor preparation of the endometrium (uterine lining) has been associated with abnormal pregnancies and a high rate of miscarriage, as well as other disorders of the female reproductive tract. Thus, the pharmaceutical compositions of the invention may be used for the support of a normal pregnancy, as well as for the treatment of abnormal pregnancies, recurring miscarriages, or the malfunction of the female reproductive tract.

Furthermore, the expression of CLH of the invention has also been found to be located in the mullerian epithel in the internal female ganglia (fallopian tubes, uterus, endocervix glands). The CLH of the invention can be used to regulate sexual differentiation, for example, by interaction with Mullerian inhibitory substances (MLS), a substance secreted by the testes, which causes the regression of the Mullerian duct system in females, leading to female sterility. In addition, the CLH of the invention may be used for the treatment of the Lawrence-Moon-Bardet-Biedl syndrome, a rare inherited condition, with variable expression, one of which is hypergenitalism (underdeveloped genitals).

In accordance with another finding of the invention, CLH was found to be expressed in tumors of the uterus, prostate and breast, indicating that CLH may be a proliferative agent on cell lines in general and tumor cell lines in particular. Thus, pharmaceutical compositions comprising an agent which decreased the expression or level of CLH, such as in anti-sense therapy, or antibodies, may be used for the treatment of these tumors.

The CLH of the invention is a hormone-responsive element, as it expressed in the mullerian epithelium, ductal epithelium of the breast, prostate, all of which are tissues under sexual hormonal control. Thus, since CLH is expressed in all estrogen target tissues (and some androgen target tissues), the pharmaceutical compositions of the invention may be used for hormonal regulation in such pathological conditions, involving non-normal amounts or a non-normal response to sexual hormones.

Pharmaceutical compositions of the invention may also be used for the treatment of cardiovascular disorders.

The nucleic acids of the invention may be used for therapeutic or diagnostic applications, for example. for the detection of the expression of CLH in various tissues, as mentioned above (for example, tumors, astrocytes, bone, tissues of the reproductive tract, etc.), and for the detection of any one of its diseases mentioned above. In addition, the ratio between the level of each of the chordin-like homologs to the other may also be indicative of a plurality of physiological or pathological conditions, for example, any one of the diseases mentioned above

The CLH gene of the invention was mapped to geonomic locus 11q14, a region containing many potential candidate bone diseases, neural system-related diseases, hormone-dependent diseases and developmental disorders. Thus, the detection of any of the CLH of the invention, as well as the detection of their amount or their ratio to each other, may be indicative to the presence of a disease, or a predisposition to a disease, or may be indicative of the severity of the disease. Furthermore, due to said association of the CLH of the invention with said disease, the pharmaceutical compositions of the invention (in connection with both aspects, i.e., both the nucleic acid sequence, the anti-sense, the amino acid sequence or the antibody) may be used for the treatment of said diseases or alleviation of some of their side effects.

The following is a list of diseases associated with the same geonomic locus as the CLH of the invention – which may be detected by the nucleic acid and amino acid sequences of the invention and the antibodies the invention and treated by the pharmaceutical compositions of the invention:

BONE RELATED DISEASES:

<u>Osteopetrosis, Autosomal Recessivea</u>

A rare hereditary disease characterized by extreme density and hardness and abnormal fragility of the bones with partial or complete obliteration of the marrow cavities. In this disorder there is a defective resorption of immature bone.

Osteoporosis-Pseudoglioma Syndrome; Oppg

A hereditary disease characterized by abnormally brittle, easily fractured bones, suggesting osteogenesis imperfecta.

High Bone Mass

High spinal bone mineral density

Osteoarthritis Susceptibility, Female-Specific

Somatotrophinoma, Acromegaly

A chronic disease of adults marked by enlargement of the bones of the extremities, face, and jaw that is caused by overactivity of the pituitary gland.

NERVOUS SYSTEM RELATED DISEASES

Pheochromocytoma, Familial Extraadrenal (Also Named Paragangliomas, Hereditary Extraadrenal)

A usually benign tumor of the adrenal medulla or the sympathetic nervous system in which the affected cells secrete increased amounts of epinephrine or norepinephrine. Disorder appears to have been due to a gene on 11q.

Tuberous Sclerosis 4

An inherited disorder of the skin and nervous system that is characterized typically by epilepsy and mental retardation, by a rash of the face resembling acne, and by multiple noncancerous tumors of the brain, kidney, retina, and heart failure, with radiographic evidence of cardiomegaly in all of them. Typical findings of tuberous sclerosis in the central nervous system, kidneys, heart, and liver.

Alexander Disease

This disorder, is characterized clinically by development of megalencephaly in infancy accompanied by progressive spasticity and dementia. In this disorder astrocytes show marked changes.

Hartnup Disorder

This disorder is characterized by a pellagra-like light-sensitive rash, cerebellar

ataxia, emotional instability, and aminoaciduria.

Spinal Muscular Atrophy With Respiratory Distress 1

Disorder that is characterized by the degeneration of motoneurons in the spinal cord resulting in muscular weakness and atrophy and that in some forms are fatal. neurogenic atrophy of skeletal muscle is observed .

Meckel Syndrome, Type 2; Mks2a

Syndrome inherited as an autosomal recessive trait and typically characterized by occipital encephalocele, microcephaly, cleft palate, polydactyly, and polycystic kidneys.

Schizophrenia Susceptibility Locus, Chromosome 11q-RELATED

Psychotic disorders usually characterized by withdrawal from reality, illogical patterns of thinking, delusions, and hallucinations, and accompanied in varying degrees by other emotional, behavioral, or intellectual disturbances. Schizophrenia, often associated with dopamine imbalances in the brain and defects of the frontal lobe, may have an underlying genetic cause.

DEVELOPMENTAL DISORDERS

Since Chordin play a role in patterning the early embryo development, Chordin-LM might involved in the following disorders:

Ebstein Anomaly

A congenital malformation of the heart that consists of downward placement of the tricuspid valve such that part of the right ventricle becomes incorporated into the pretricuspid chamber. Rearrangements of the long arm of chromosome 11 were described in patients with Ebstein anomaly.

Rutledge Lethal Multiple Congenital Anomaly Syndrome

External features, mesomelic dwarfism, micrognathia, V-shaped upper lip, microglossia, thick alveolar ridges, ambiguous genitalia, webbed neck, highly arched palate, clubfeet, fused fontanelles, inclusion cysts of the tongue, widely spaced nipples, and digital anomalies. Internal findings included oligopapillary renal hypoplasia, severe congenital heart defect, cerebellar hypoplasia, and

pulmonary, laryngeal, and gallbladder hypoplasia.

Bardet-Biedl Syndrome, Type 1; Bbs1

The Bardet-Biedl syndrome is characterized by mental retardation, pigmentary retinopathy, polydactyly, obesity, and hypogenitalism. The disorder is inherited as an autosomal recessive.

Targeted inactivation of chordin results in animals that display defects in inner and outer ear development. Therefore chordin-LM might be involved in hearing disorders such as the one linked to chromosome 11 - DEAFNESS, Autosomal Dominant Nonsyndromic Sensorineural 11; Dfna11.

The present invention also provides expression vectors comprising any one of the above defined complementary nucleic acid sequences and host cells transfected with said nucleic acid sequences or vectors, being complementary to those specified in the first aspect of the invention.

The invention also provides anti-CLH product antibodies, namely antibodies directed against the CLH product which specifically bind to said CLH product. Said antibodies are useful both for diagnostic and therapeutic purposes. For example said antibody may be as an active ingredient in a pharmaceutical composition as will be explained below.

The present invention also provides pharmaceutical compositions comprising, as an active ingredient, the nucleic acid molecules which comprise or consist of said complementary sequences, or of a vector comprising said complementary sequences. The pharmaceutical composition thus provides pharmaceutical compositions comprising, as an active ingredient, said anti-CLH product antibodies.

The pharmaceutical compositions comprising said anti-CLH product antibodies or the nucleic acid molecule comprising said complementary sequence, are suitable for the treatment of diseases and pathological conditions where a therapeutically beneficial effect may be achieved by neutralizing the CLH or decreasing the amount of the CLH product or blocking its binding to its target (for example its receptor), for example, by the neutralizing effect of the antibodies, or

by the decrease of the effect of the antisense mRNA in decreasing expression level of the CLH product. Examples of the diseases are any one of those mentioned above.

According to the third aspect of the invention the present invention provides methods for detecting the level of the transcript (mRNA) of said CLH product in a body fluid sample, or in a specific tissue sample or body fluid, for example, by use of probes comprising or consisting of said coding sequences (or complementary sequences); as well as methods for detecting levels of expression of said product in tissue, e.g. by the use of antibodies capable of specifically reacting with the above amino acid sequences. Detection of the level of the expression of the CLH of the invention may be indicative of a plurality of physiological or pathological conditions.

The method, according to this latter aspect, for detection of a nucleic acid sequence which encodes the CLH product in a biological sample, comprises the steps of:

(a) providing a probe comprising at least one of the nucleic acid sequence defined above;

(b) contacting the biological sample with said probe under conditions allowing hybridization of nucleic acid sequences thereby enabling formation of hybridization complexes;

(c) detecting hybridization complexes, wherein the presence of the complex indicates the presence of nucleic acid sequence encoding the CLH product in the biological sample.

The amount of hybridization complexes may be determined and calibrated by comparing it to a calibration scale in order to determine the amount of the nucleic acid sequence which enables the CLH product in the sample. The level of each of the sequences may be detected and either compared to the calibrated levels or to the level of each other, and said ratio may also be indicative to a plurality of pathological or physiological conditions.

By a preferred embodiment the probe is part of a nucleic acid chip used for detection purposes, i.e. the probe is a part of an array of probes each present in a known location on a solid support.

The nucleic acid sequence used in the above method may be a DNA sequence an RNA sequence, etc; it may be a coding or a sequence or a sequence complementary thereto (for respective detection of RNA transcripts or coding-DNA sequences). By quantization of the level of hybridization complexes and calibrating the quantified results it is possible also to detect the level of the transcript in the sample.

Methods for detecting mutations in the region coding for the CLH product are also provided, which may be methods carried-out in a binary fashion, namely merely detecting whether there is any mismatches between the normal CLH nucleic acid sequence and the one present in the sample, or carried-out by specifically detecting the nature and location of the mutation.

The present invention also concerns a method for detecting CLH product both for determining its presence, as well as its level or alterations in its level in a biological sample, comprising the steps of:

(a) contacting with said biological sample the antibody of the invention, thereby forming an antibody-antigen complex; and

(b) detecting said antibody-antigen complex

wherein the presence of said antibody-antigen complex correlates with the presence of CLH product in said biological sample.

The present invention also concerns a method for detecting anti-CLH antibodies in a biological sample comprising the steps of:

(a) contacting said biological sample with the product of the invention thereby forming an antibody-antigen complex; and

(b) detecting said antibody-antigen complex

wherein the presence of said antibody-antigen complex correlates with the presence of anti-CHL antibody in said biological sample.

In many cases, diseases are detected not by detecting the presence of the protein (product) which caused the disease, but rather by detecting the presence in a biological sample (such as blood or serum) of antibodies against such a product.

The method of detecting the presence of anti-CLH antibodies is intended to be used in such case.

The amount of the antibody-antigen complex can be quantitized, in order to determine the level of the CHL-product or the anti-CHL antibodies, as the case may be.

As explained above, the level of any one of the products may be compared to each other, and the ratio between the levels may be indicative to a plurality of physiological and pathological conditions. In addition, the indicative ratio may not be the ratio of the proteins themselves but rather the ratio of antibodies against the proteins.

By yet another aspect the invention also provides a method for identifying candidate compounds capable of modulating the activity of CLH product (being either activators or deactivators). The method includes:

(i) providing a protein or polypeptide comprising an amino acid sequence substantially as depicted in any one of SEQ ID NO: 12 to SEQ ID NO: 22, or a fragment of such a sequence;

(ii) contacting a candidate compound with said amino acid sequence;

(iii) comprising the physiological effect of the amino acid sequence in the presence and absence of said candidate compound and selecting those compounds which show a significant effect on said physiological activity.

The activity of the amino acid which should be changed by the modulator (being either the activator or deactivator) may be for example the binding of the CLH product to its receptor, the effect of CLH on BMPs expression or activity. Any modulator which changes such an activity has an infecting potential, as serving as an actuator or deactivator.

The present invention also concerns compounds identified by the above methods described above, which compound may either be an activator of the CLH product or a deactivator thereof.

BRIEF DESCRIPTION OF THE DRAWINGS

In order to understand the invention and to see how it may be carried out in practice, a preferred embodiment will now be described, by way of non-limiting example only, with reference to the accompanying drawings, in which:

Fig. 1 is alignment of the CLH product of SEQ ID NO: 12 to known chordin protein, demonstrating the homology regions within these proteins. The alignment was performed using best-fit of GCG;

Fig. 2a is the alignment of the first splice variant (SEQ ID NO: 13) to the known chordin deposited in the Emb as gi 4808227;

Fig. 2b is the alignment of the first splice variant (SEQ ID NO: 13) to the known chordin deposited in the Emb under gi 3822218;

Fig. 2c is the alignment of the first splice variant (SEQ ID NO: 13) to the known chordin deposited in the Emb under gi 3800772;

Fig. 3a is the alignment of the second splice variant (SEQ ID NO: 14) with a known chordin deposited in the Emb under gi 4808227;

Fig. 3b is the alignment of the second splice variant (SEQ ID NO: 14) with a known chordin deposited in the Emb under gi 3822218;

Fig. 4a is the alignment of the third splice variant (SEQ ID NO: 15) with a known chordin deposited in the Emb under gi 4808227;

Fig. 4b is the alignment of the third splice variant (SEQ ID NO: 15) with a known chordin deposited in the Emb under gi 2731578;

Fig. 4c is the alignment of the third splice variant (SEQ ID NO: 16) with a known chordin deposited in the Emb under gi 2498235;

Fig. 4d is the alignment of the third splice variant (SEQ ID NO: 16) with a known chordin deposited in the Emb under gi 3822218;

Fig. 5 is multiple alignments of the sequences of the first four splice variants to several known chordins;.

Fig. 6 is the alignment of SEQ ID No. 16 to the known chordin deposited as gi 48082227;

Fig. 7 is the alignment of SEQ ID No. 16 to the known chordin deposited as gi 3822218;

Fig. 8 is the alignment of SEQ ID No. 16 to the known chordin deposited as gi 6753418;

Fig. 9 shows the alignment of SEQ ID No. 17 to the known chording deposited as gi 4808227;

Fig. 10 is the alignment of SEQ ID No. 17 to the known chordin deposited as gi 3822218;

Fig. 11 shows the alignment of SEQ ID No. 18 to the known chordin deposited as gi 4808222;.

Fig. 12 shows the alignment of SEQ ID No. 18 to the known chordin deposited as gi 3822218;

Fig. 13 shows the alignment of SEQ ID No. 19 to the known chordin deposited as gi 2731578;

Fig. 14 shows the alignment of SEQ ID No. 18 to the known chordin deposited as gi 3822218;

Fig. 15 shows the alignment of SEQ ID No. 20 to the known chordin deposited as gi 2731578;

Fig. 16 shows the alignment of SEQ ID No. 20 to the known chordin deposited as gi 382218;

Fig. 17 shows the alignment of SEQ ID No. 21 to the known chordin deposited as gi 2731578;

Fig. 18 shows the alignment of SEQ ID No. 21 to the the known chordin deposited as gi 3822218.

Fig. 19 shows multiple alignments of SEQ ID Nos. 12-21 (termed var 1-var 6, respectively) to each other;

Fig. 20 shows the alignment of SEQ ID No. 22 to the known chordin deposited as gi 480827.

Fig. 21 shows the alignment of SEQ ID No. 22 to the known chordin deposited as gi 6753418.

Fig. 22 shows a Northern blot analysis of CLH expression in: skeletal muscles, uterus, colon, small intestine, bladder, heart, stomach, prostate;

Fig. 23 shows a Western blot analysis of transfected COS-7 cells which express and secrete CHL;

Fig. 24 shows immunohistochemistry results with breast carcinoma (ductal and invasive ductal); prostate (carcinoma and benign prostate hyperplasia); bladder transitional epithelium; Mullerian Epithelium; uterus, bone Glioblastoma Multi-form (GBM); and

Fig. 25 shows Western blot analysis, or expression of CLH in human brain and bone tissues;

DETAILED DESCRIPTION OF A PREFERRED EMBODIMENT

Example I: CLH - nucleic acid sequence

The nucleic acid sequences of the invention include nucleic acid sequences which encode CLH product and fragments and analogs thereof. The nucleic acid sequences may alternatively be sequences complementary to the above coding sequence, or to a region of said coding sequence. The length of the complementary sequence is sufficient to avoid the expression of the coding sequence. The nucleic acid sequences may be in the form of RNA or in the form of DNA, and include messenger RNA, synthetic RNA and DNA, cDNA, and genomic DNA. The DNA may be double-stranded or single-stranded, and if single-stranded may be the coding strand or the non-coding (anti-sense, complementary) strand. The nucleic acid sequences may also both include dNTPs, rNTPs as well as non naturally occurring sequences. The sequence may also be a part of a hybrid between an amino acid sequence and a nucleic acid sequence.

In a general embodiment, the nucleic acid sequence has at least 70%, preferably 80% or 90% or 95% sequence identity with any one of the sequences identified as SEQ ID NO: 1 to SEQ ID NO: 11.

The nucleic acid sequences may include the coding sequence by itself. By another alternative the coding region may be in combination with additional coding sequences, such as those coding for fusion protein or signal peptides, in combination with non-coding sequences, such as introns and control elements, promoter and terminator elements or 5' and/or 3' untranslated regions, effective for expression of the coding sequence in a suitable host, and/or in a vector or host environment in which the CLH nucleic acid sequence is introduced as a heterologous sequence.

The nucleic acid sequences of the present invention may also have the product coding sequence fused in-frame to a marker sequence which allows for purification of the CLH product. The marker sequence may be, for example, a

hexahistidine tag to provide for purification of the mature polypeptide fused to the marker in the case of a bacterial host, or, the marker sequence may be a hemagglutinin (HA) tag when a mammalian host, e.g. COS-7 cells, is used. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein (Wilson, I., *et al. Cell* 37:767 (1984)).

Also included in the scope of the invention are fragments also referred to herein as oligonucleotides, typically having at least 20 bases, preferably 20-30 bases corresponding to a region of the coding-sequence nucleic acid sequence. The fragments may be used as probes, primers, and when complementary also as antisense agents, and the like, according to known methods.

As indicated above, the nucleic acid sequence may be substantially a depicted in any one of SEQ ID NO: 1 to SEQ ID NO: 11 or fragments thereof or sequences having at least 70%, preferably 70-80%, most preferably 90% or 95% identity to the above sequence. Alternatively, due to the degenerative nature of the genetic code, the sequence may be a sequence coding the amino acid sequence of any one of SEQ ID NO: 12 to SEQ ID NO: 22, or fragments or analogs of said amino acid sequence.

A. Preparation of nucleic acid sequences

The nucleic acid sequences may be obtained by screening cDNA libraries using oligonucleotide probes which can hybridize to or PCR-amplify nucleic acid sequences which encode the CLH products disclosed above. cDNA libraries prepared from a variety of tissues are commercially available and procedures for screening and isolating cDNA clones are well-known to those of skill in the art. Such techniques are described in, for example, Sambrook *et al.* (1989) Molecular Cloning: A Laboratory Manual (2nd Edition), Cold Spring Harbor Press, Plainview, N.Y. and Ausubel FM et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y.

The nucleic acid sequences may be extended to obtain upstream and downstream sequences such as promoters, regulatory elements, and 5' and 3'

untranslated regions (UTRs). Extension of the available transcript sequence may be performed by numerous methods known to those of skill in the art, such as PCR or primer extension (Sambrook *et al., supra*), or by the RACE method using, for example, the Marathon RACE kit (Clontech, Cat. # K1802-1).

Alternatively, the technique of *"restriction-site"* PCR (Gobinda *et al. PCR Methods Applic.* **2**:318-22, (1993)), which uses universal primers to retrieve flanking sequence adjacent a known locus, may be employed. First, genomic DNA is amplified in the presence of primer to a linker sequence and a primer specific to the known region. The amplified sequences are subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR can be used to amplify or extend sequences using divergent primers based on a known region (Triglia, T. *et al.*, *Nucleic Acids Res.* **16**:8186, (1988)). The primers may be designed using OLIGO(R) 4.06 Primer Analysis Software (1992; National Biosciences Inc, Plymouth, Minn.), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68-72°C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Capture PCR (Lagerstrom, M. *et al., PCR Methods Applic.* **1**:111-19, (1991)) is a method for PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA. Capture PCR also requires multiple restriction enzyme digestions and ligations to place an engineered double-stranded sequence into a flanking part of the DNA molecule before PCR.

Another method which may be used to retrieve flanking sequences is that of Parker, J.D., *et al., Nucleic Acids Res.*, **19**:3055-60, (1991)). Additionally, one can use PCR, nested primers and PromoterFinder™ libraries to *"walk in"* genomic

DNA (PromoterFinder™; Clontech, Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions. Preferred libraries for screening for full length cDNAs are ones that have been size-selected to include larger cDNAs. Also, random primed libraries are preferred in that they will contain more sequences which contain the 5' and upstream regions of genes.

A randomly primed library may be particularly useful if an oligo d(T) library does not yield a full-length cDNA. Genomic libraries are useful for extension into the 5' nontranslated regulatory region.

The nucleic acid sequences and oligonucleotides of the invention can also be prepared by solid-phase methods, according to known synthetic methods. Typically, fragments of up to about 100 bases are individually synthesized, then joined to form continuous sequences up to several hundred bases.

B. Use of CLH nucleic acid sequence for the production of CLH products

In accordance with the present invention, nucleic acid sequences specified above may be used as recombinant DNA molecules that direct the expression of CLH products.

As will be understood by those of skill in the art, it may be advantageous to produce CLH product-encoding nucleotide sequences possessing codons other than those which appear in any one of SEQ ID NO: 1 to SEQ ID NO: 11 which are those which naturally occur in the human genome. Codons preferred by a particular prokaryotic or eukaryotic LHost (Murray, E. *et al. Nuc Acids Res.*, **17**:477-508, (1989)) can be selected, for example, to increase the rate of CLH product expression or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, than transcripts produced from naturally occurring sequence.

The nucleic acid sequences of the present invention can be engineered in order to alter a CLH product coding sequence for a variety of reasons, including but not limited to, alterations which modify the cloning, processing and/or

expression of the product. For example, alterations may be introduced using techniques which are well known in the art, e.g., site-directed mutagenesis, to insert new restriction sites, to alter glycosylation patterns, to change codon preference, to produce splice variants, etc.

The present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a nucleic acid sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are also described in Sambrook, *et al., (supra)*.

The present invention also relates to host cells which are genetically engineered with vectors of the invention, and the production of the product of the invention by recombinant techniques. Host cells are genetically engineered (i.e., transduced, transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the expression of the CLH nucleic acid sequence. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those skilled in the art.

The nucleic acid sequences of the present invention may be included in any one of a variety of expression vectors for expressing a product. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral

DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host. The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and related sub-cloning procedures are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate transcription control sequence (promoter) to direct mRNA synthesis. Examples of such promoters include: LTR or SV40 promoter, the *E.coli lac* or *trp* promoter, the phage lambda *PL* promoter, and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation, and a transcription terminator. The vector may also include appropriate sequences for amplifying expression. In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E.coli*.

The vector containing the appropriate DNA sequence as described above, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein. Examples of appropriate expression hosts include: bacterial cells, such as *E.coli*, *Streptomyces, Salmonella typhimurium*; fungal cells, such as yeast; insect cells such as *Drosophila* and *Spodoptera* Sf9; animal cells such as CHO, COS, HEK 293 or Bowes melanoma; adenoviruses; plant cells, etc. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein. The invention is not limited by the host cells employed.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for the CLH product. For example, when large

quantities of CLH product are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be desirable. Such vectors include, but are not limited to, multifunctional *E.coli* cloning and expression vectors such as *Bluescript*(R) (Stratagene), in which the CLH polypeptide coding sequence may be ligated into the vector in-frame with sequences for the amino-terminal Met and the subsequent 7 residues of beta-galactosidase so that a hybrid protein is produced; *pIN* vectors (Van Heeke & Schuster *J. Biol. Chem.* **264**:5503-5509, (1989)); *pET* vectors (Novagen, Madison WI); and the like.

In the yeast *Saccharomyces cerevisiae* a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase and PGH may be used. For reviews, see Ausubel *et al. (supra)* and Grant *et al.*, (*Methods in Enzymology* **153**:516-544, (1987)).

In cases where plant expression vectors are used, the expression of a sequence encoding CLH product may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of *CaMV* (Brisson *et al., Nature* **310**:511-514. (1984)) may be used alone or in combination with the omega leader sequence from TMV (Takamatsu *et al.*, *EMBO J.*, **6**:307-311, (1987)). Alternatively, plant promoters such as the small subunit of RUBISCO (Coruzzi *et al.*, *EMBO J.* **3**:1671-1680, (1984); Broglie *et al.*, *Science* **224**:838-843, (1984)); or heat shock promoters (Winter J and Sinibaldi R.M., *Results Probl. Cell Differ.*, **17**:85-105, (1991)) may be used. These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. For reviews of such techniques, see Hobbs S. or Murry L.E. (1992) in McGraw Hill Yearbook of Science and Technology, McGraw Hill, New York, N.Y., pp 191-196; or Weissbach and Weissbach (1988) *Methods for Plant Molecular Biology*, Academic Press, New York, N.Y., pp 421-463.

CLH product may also be expressed in an insect system. In one such system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a

vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. The CLH product coding sequence may be cloned into a nonessential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of CLH coding sequence will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein coat. The recombinant viruses are then used to infect *S. frugiperda* cells or *Trichoplusia* larvae in which CLH protein is expressed (Smith *et al., J. Virol.* 46:584, (1983); Engelhard, E.K. *et al., Proc. Nat. Acad. Sci.* **91**:3224-7, (1994)).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, a CLH product coding sequence may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a nonessential E1 or E3 region of the viral genome will result in a viable virus capable of expressing CLH protein in infected host cells (Logan and Shenk, *Proc. Natl. Acad. Sci.* **81**:3655-59, (1984). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be required for efficient translation of a CLH protein coding sequence. These signals include the ATG initiation codon and adjacent sequences. In cases where CLH product coding sequence, its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous transcriptional control signals including the ATG initiation codon must be provided. Furthermore, the initiation codon must be in the correct reading frame to ensure transcription of the entire insert. Exogenous transcriptional elements and initiation codons can be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate to the cell system in use (Scharf, D. *et al.*,

(1994) *Results Probl. Cell Differ.*, **20**:125-62, (1994); Bittner et al., *Methods in Enzymol* **153**:516-544, (1987)).

In a further embodiment, the present invention relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (Davis, L., Dibner, M., and Battey, I. (1986) Basic Methods in Molecular Biology). Cell-free translation systems can also be employed to produce polypeptides using RNAs derived from the DNA constructs of the present invention.

A host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the protein include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation. Post-translational processing which cleaves a *"pre-pro"* form of the protein may also be important for correct insertion, folding and/or function. Different host cells such as CHO, HeLa, MDCK, 293, WI38, etc. have specific cellular machinery and characteristic mechanisms for such post-translational activities and may be chosen to ensure the correct modification and processing of the introduced, foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express CLH product may be transformed using expression vectors which contain viral origins of replication or endogenous expression elements and a selectable marker gene. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the

introduced sequences. Resistant clumps of stably transformed cells can be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler M., *et al.*, *Cell* **11**:223-32, (1977)) and adenine phosphoribosyltransferase (Lowy I., *et al.*, *Cell* **22**:817-23, (1980)) genes which can be employed in *tk-* or *aprt-* cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, *dhfr* which confers resistance to methotrexate (Wigler M., *et al., Proc. Natl. Acad. Sci.* **77**:3567-70, (1980)); *npt*, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. *et al., J. Mol. Biol.*, **150**:1-14, (1981)) and *als* or *pat*, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, *supra*). Additional selectable genes have been described, for example, *trpB*, which allows cells to utilize indole in place of tryptophan, or *hisD*, which allows cells to utilize histinol in place of histidine (Hartman S.C. and R.C. Mulligan, *Proc. Natl. Acad. Sci.* **85**:8047-51, (1988)). The use of visible markers has gained popularity with such markers as anthocyanins, beta-glucuronidase and its substrate, GUS, and luciferase and its substrates, luciferin and ATP, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. *et. al., Methods Mol. Biol.*, 55:121-131, (1995)).

Host cells transformed with a nucleotide sequence encoding CLH product may be cultured under conditions suitable for the expression and recovery of the encoded protein from cell culture. The product produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing nucleic acid sequences encoding CLH product can be designed with signal sequences which direct secretion of CLH product through a prokaryotic or eukaryotic cell membrane.

CLH product may also be expressed as a recombinant protein with one or more additional polypeptide domains added to facilitate protein purification. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp, Seattle, Wash.). The inclusion of a protease-cleavable polypeptide linker sequence between the purification domain and CLH protein is useful to facilitate purification. One such expression vector provides for expression of a fusion protein compromising a CLH polypeptide fused to a polyhistidine region separated by an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography, as described in Porath, *et al.*, *Protein Expression and Purification,* **3**:263-281, (1992)) while the enterokinase cleavage site provides a means for isolating CLH polypeptide from the fusion protein. *pGEX* vectors (Promega, Madison, Wis.) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to ligand-agarose beads (e.g., glutathione-agarose in the case of GST-fusions) followed by elution in the presence of free ligand.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, or other methods, which are well know to those skilled in the art.

The CLH products can be recovered and purified from recombinant cell cultures by any of a number of methods well known in the art, including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

C. Diagnostic applications utilizing nucleic acid sequences

The nucleic acid sequences of the present invention may be used for a variety of diagnostic purposes. The nucleic acid sequences may be used to detect and quantitate expression of CLH in patient's cells, e.g. biopsied tissues, by detecting the presence of mRNA coding for CLH product. Alternatively, the assay may be used to detect soluble CLH in the serum or blood. This assay typically involves obtaining total mRNA from the tissue or serum and contacting the mRNA with a nucleic acid probe. The probe is a nucleic acid molecule of at least 20 nucleotides, preferably 20-30 nucleotides, capable of specifically hybridizing with a sequence included within the sequence of a nucleic acid molecule encoding CLH under hybridizing conditions, detecting the presence of mRNA hybridized to the probe, and thereby detecting the expression of CLH. This assay can be used to distinguish between absence, presence, and excess expression of CLH product and to monitor levels of CLH expression during therapeutic intervention.

The invention also contemplates the use of the nucleic acid sequences as a diagnostic for diseases resulting from inherited defective CLH sequences. These sequences can be detected by comparing the sequences of the defective (i.e., mutant) CLH coding region with that of a normal coding region. Association of the sequence coding for mutant CLH product with abnormal CLH product

activity may be verified. In addition, sequences encoding mutant CLH products can be inserted into a suitable vector for expression in a functional assay system (e.g., colorimetric assay, complementation experiments in a CLH protein deficient strain of HEK293 cells) as yet another means to verify or identify mutations. Once mutant genes have been identified, one can then screen populations of interest for carriers of the mutant gene.

Individuals carrying mutations in the nucleic acid sequence of the present invention may be detected at the DNA level by a variety of techniques. Nucleic acids used for diagnosis may be obtained from a patient's cells, including but not limited to such as from blood, urine, saliva, placenta, tissue biopsy and autopsy material. Genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR (Saiki, *et al.*, *Nature* **324**:163-166, (1986)) prior to analysis. RNA or cDNA may also be used for the same purpose. As an example, PCR primers complementary to the nucleic acid of the present invention can be used to identify and analyze mutations in the gene of the present invention. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype.

Point mutations can be identified by hybridizing amplified DNA to radiolabeled RNA of the invention or alternatively, radiolabeled antisense DNA sequences of the invention. Sequence changes at specific locations may also be revealed by nuclease protection assays, such RNase and S1 protection or the chemical cleavage method (e.g. Cotton, *et alProc. Natl. Acad. Sci.* USA, **85**:4397-4401, (1985)), or by differences in melting temperatures. *"Molecular beacons"* (Kostrikis L.G. *et al.*, Science **279**:1228-1229, (1998)), hairpin-shaped, single-stranded synthetic oligo- nucleotides containing probe sequences which are complementary to the nucleic acid of the present invention, may also be used to detect point mutations or other sequence changes as well as monitor expression levels of CLH product. Such diagnostics would be particularly useful for prenatal testing.

Another method for detecting mutations uses two DNA probes which are designed to hybridize to adjacent regions of a target, with abutting bases, where the region of known or suspected mutation(s) is at or near the abutting bases. The two probes may be joined at the abutting bases, e.g., in the presence of a ligase enzyme, but only if both probes are correctly base paired in the region of probe junction. The presence or absence of mutations is then detectable by the presence or absence of ligated probe.

Also suitable for detecting mutations in the CLH product coding sequence are oligonucleotide array methods based on sequencing by hybridization (SBH), as described, for example, in U.S. Patent No. 5,547,839. In a typical method, the DNA target analyte is hybridized with an array of oligonucleotides formed on a microchip. The sequence of the target can then be *"read"* from the pattern of target binding to the array.

D. Gene mapping utilizing nucleic acid sequences

The nucleic acid sequences of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. Moreover, there is a current need for identifying particular sites on the chromosome. Few chromosome marking reagents based on actual sequence data (repeat polymorphisms) are presently available for marking chromosomal location. The mapping of DNAs to chromosomes according to the present invention is an important first step in correlating those sequences with genes associated with disease.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 20-30 bp) from the CLH cDNA. Computer analysis of the 3' untranslated region is used to rapidly select primers that do not span more than one exon in the genomic DNA, which would complicate the amplification process. These primers are then used for PCR screening of somatic cell hybrids

containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the primer will yield an amplified fragment.

PCR mapping of somatic cell hybrids or using instead radiation hybrids are rapid procedures for assigning a particular DNA to a particular chromosome. Using the present invention with the same oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes or pools of large genomic clones in an analogous manner. Other mapping strategies that can similarly be used to map to its chromosome include *in situ* hybridization, prescreening with labeled flow-sorted chromosomes and preselection by hybridization to construct chromosome specific-cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a cDNA clone to a metaphase chromosomal spread can be used to provide a precise chromosomal location in one step. This technique can be used with cDNA as short as 50 or 60 bases. For a review of this technique, see Verma *et al., Human Chromosomes: a Manual of Basic Techniques*, (1988) Pergamon Press, New York.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, for example, in the OMIM database (Center for Medical Genetics, Johns Hopkins University, Baltimore, MD and National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD). The OMIM gene map presents the cytogenetic map location of disease genes and other expressed genes. The OMIM database provides information on diseases associated with the chromosomal location. Such associations include the results of linkage analysis mapped to this interval, and the correlation of translocations and other chromosomal aberrations in this area with the advent of various diseases, for example, those mentioned in connection with the pharmaceutical compositions of the invention.

E. Therapeutic applications of nucleic acid sequences

Nucleic acid sequences of the invention may also be used for therapeutic purposes. Turning first to the second aspect of the invention (i.e. inhibition of expression of CLH), expression of CLH product may be modulated through antisense technology, which controls gene expression through hybridization of complementary nucleic acid sequences, i.e. antisense DNA or RNA, to the control, 5' or regulatory regions of the gene encoding CLH product. For example, the 5' coding portion of the nucleic acid sequence sequence which codes for the product of the present invention is used to design an antisense oligonucleotide of from about 10 to 40 base pairs in length. Oligonucleotides derived from the transcription CLHt site, e.g. between positions -10 and +10 from the CLHt site, are preferred. An antisense DNA oligonucleotide is designed to be complementary to a region of the nucleic acid sequence involved in transcription (Lee *et al., Nucl. Acids, Res.*, **6**:3073, (1979); Cooney et al., Science **241**:456, (1988); and Dervan *et al., Science* **251**:1360, (1991)), thereby preventing transcription and the production of the CLH products. An antisense RNA oligonucleotide hybridizes to the mRNA in vivo and blocks translation of the mRNA molecule into the CLH products (Okano *J. Neurochem.* **56**:560, (1991)). The antisense constructs can be delivered to cells by procedures known in the art such that the antisense RNA or DNA may be expressed *in vivo*. The antisense may be antisense mRNA or DNA sequence capable of coding such antisense mRNA. The antisense mRNA or the DNA coding thereof can be complementary to the full sequence of nucleic acid sequences coding to the CLH protein or to a fragment of such a sequence which is sufficient to inhibit production of a protein product.

Turning now to the first aspect of the invention, i.e. expression of CLH, expression of CLH product may be increased by providing coding sequences for coding for said product under the control of suitable control elements ending its expression in the desired host.

The nucleic acid sequences of the invention may be employed in combination with a suitable pharmaceutical carrier. Such compositions comprise a therapeutically effective amount of the compound, and a pharmaceutically acceptable carrier or excipient. Such a carrier includes but is not limited to saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The formulation should suit the mode of administration.

The products of the invention as well as any activators and deactivators compounds (see below) which are polypeptides, may also be employed in accordance with the present invention by expression of such polypeptides *in vivo*, which is often referred to as *"gene therapy."* Cells from a patient may be engineered with a nucleic acid sequence (DNA or RNA) encoding a polypeptide *ex vivo*, with the engineered cells then being provided to a patient to be treated with the polypeptide. Such methods are well-known in the art. For example, cells may be engineered by procedures known in the art by use of a retroviral particle containing RNA encoding a polypeptide of the present invention.

Similarly, cells may be engineered in vivo for expression of a polypeptide *in vivo* by procedures known in the art. As known in the art, a producer cell for producing a retroviral particle containing RNA encoding the polypeptide of the present invention may be administered to a patient for engineering cells *in vivo* and expression of the polypeptide *in vivo*. These and other methods for administering a product of the present invention by such method should be apparent to those skilled in the art from the teachings of the present invention. For example, the expression vehicle for engineering cells may be other than a retrovirus, for example, an adenovirus which may be used to engineer cells *in vivo* after combination with a suitable delivery vehicle.

Retroviruses from which the retroviral plasmid vectors mentioned above may be derived include, but are not limited to, Moloney Murine Leukemia Virus, spleen necrosis virus, retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, gibbon ape leukemia virus, human immunodeficiency virus, adenovirus, Myeloproliferative Sarcoma Virus, and mammary tumor virus.

The retroviral plasmid vector is employed to transduce packaging cell lines to form producer cell lines. Examples of packaging cells which may be transfected include, but are not limited to, the *PE501*, *PA317*, *psi-2*, *psi-AM*, *PA12*, *T19-14X*, *VT-19-17-H2*, *psi-CRE*, *psi-CRIP*, *GP+E-86*, *GP+envAm12*, and *DAN* cell lines as described in Miller (*Human Gene Therapy*, Vol. **1**, pg. 5-14, (1990)). The vector may transduce the packaging cells through any means known in the art. Such means include, but are not limited to, electroporation, the use of liposomes, and $CaPO_4$ precipitation. In one alternative, the retroviral plasmid vector may be encapsulated into a liposome, or coupled to a lipid, and then administered to a host.

The producer cell line generates infectious retroviral vector particles which include the nucleic acid sequence(s) encoding the polypeptides. Such retroviral vector particles then may be employed, to transduce eukaryotic cells, either in vitro or in vivo. The transduced eukaryotic cells will express the nucleic acid sequence(s) encoding the polypeptide. Eukaryotic cells which may be transduced include, but are not limited to, embryonic stem cells, embryonic carcinoma cells, as well as hematopoietic stem cells, hepatocytes, fibroblasts, myoblasts, keratinocytes, endothelial cells, and bronchial epithelial cells.

The genes introduced into cells may be placed under the control of inducible promoters, such as the radiation-inducible Egr-1 promoter, (Maceri, H.J., *et al., Cancer Res.*, **56**(19):4311 (1996)), to stimulate CLH production or antisense inhibition in response to radiation, eg., radiation therapy for treating tumors.

Example II. CLH product

The substantially purified CLH product of the invention has been defined above as the product coded from the nucleic acid sequence of the invention. Preferably the amino acid sequence is an amino acid sequence having at least 70%, preferably at least 80% or 90% or 95% identity to the sequence identified as any one of SEQ ID NO: 12 to SEQ ID NO: 22. The protein or polypeptide may

be in mature and/or modified form, also as defined above. Also contemplated are protein fragments having at least 10 contiguous amino acid residues, preferably at least 10-20 residues, derived from the CLH product.

The sequence variations are preferably those that are considered conserved substitutions, as defined above. Thus, for example, a protein with a sequence having at least 80%, preferably 90% sequence identity with the protein identified as any one of SEQ ID NO: 12 to SEQ ID NO: 22, preferably by utilizing conserved substitutions as defined above is also part of the invention. In a more specific embodiment, the protein has or contains the sequence identified as any one of SEQ ID NO: 12 to SEQ ID NO: 22. The CLH product may be (i) one in which one or more of the amino acid residues in a sequence listed above are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue), or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the CLH product is fused with another compound, such as a compound to increase the half-life of the protein (for example, polyethylene glycol (PEG)), or a moiety which serves as targeting means to direct the protein to its target tissue or target cell population (such as an antibody), or (iv) one in which additional amino acids are fused to the CLH product. Such fragments, variants and derivatives are deemed to be within the scope of those skilled in the art from the teachings herein.

A. Preparation of CLH product

Recombinant methods for producing and isolating the CLH product, and fragments of the protein are described above.

In addition to recombinant production, fragments and portions of CLH product may be produced by direct peptide synthesis using solid-phase techniques (cf. Stewart *et al.*, (1969) Solid-Phase Peptide Synthesis, WH Freeman Co, San Francisco; Merrifield J., *J. Am. Chem. Soc.*, **85**:2149-2154, (1963)). In vitro peptide synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems

431A Peptide Synthesizer (Perkin Elmer, Foster City, Calif.) in accordance with the instructions provided by the manufacturer. Fragments of CLH product may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

B. Therapeutic uses and compositions utilizing the CLH product

The CLH product of the invention is generally useful in treating diseases and disorders which are characterized by a lower than normal level of CLH expression, and or diseases which can be cured or ameliorated by raising the level of the CLH product, even if the level is normal.

Typically these diseases are in CLH products or fragments and may be administered by any of a number of routes and methods designed to provide a consistent and predictable concentration of compound at the target organ or tissue. The product-containing compositions may be administered alone or in combination with other agents, such as stabilizing compounds, and/or in combination with other pharmaceutical agents such as drugs or hormones.

CLH product-containing compositions may be administered by a number of routes including, but not limited to oral, intravenous, intramuscular, transdermal, subcutaneous, topical, sublingual, or rectal means as well as by nasal application. CLH product-containing compositions may also be administered via liposomes. Such administration routes and appropriate formulations are generally known to those of skill in the art.

The product can be given via intravenous or intraperitoneal injection. Similarly, the product may be injected to other localized regions of the body. The product may also be administered via nasal insufflation. Enteral administration is also possible. For such administration, the product should be formulated into an appropriate capsule or elixir for oral administration, or into a suppository for rectal administration.

The foregoing exemplary administration modes will likely require that the product be formulated into an appropriate carrier, including ointments, gels,

suppositories. Appropriate formulations are well known to persons skilled in the art.

Dosage of the product will vary, depending upon the potency and therapeutic index of the particular polypeptide selected.

A therapeutic composition for use in the treatment method can include the product in a sterile injectable solution, the polypeptide in an oral delivery vehicle, the product in an aerosol suitable for nasal administration, or the product in a nebulized form, all prepared according to well known methods. Such compositions comprise a therapeutically effective amount of the compound, and a pharmaceutically acceptable carrier or excipient. Such a carrier includes but is not limited to saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The product of the invention may also be used to modulate endothelial differentiation and proliferation as well as to modulate apoptosis either *ex vivo* or *in vitro*, for example, in cell cultures.

Example III. Screening methods for activators and deactivators (inhibitors)

The present invention also includes an assay for identifying molecules, such as synthetic drugs, antibodies, peptides, or other molecules, which have a modulating effect on the activity of the CLH product, e.g. activators or deactivators of the CLH product of the present invention. Such an assay comprises the steps of providing an CLH product encoded by the nucleic acid sequences of the present invention and determining its physiological activity on the target in the presence and absence of one or more candidate molecules to determine the candidate molecules. Those molecules which are modulating effect on the activity of the CLH product are selected as likely candidates for activators and deactivators.

CLH product, its catalytic or immunogenic fragments or oligopeptides thereof, can be used for screening therapeutic compounds in any of a variety of drug screening techniques. The fragment employed in such a test may be free in solution, affixed to a solid support, borne on a cell membrane or located

intracellularly. The formation of binding complexes, between CLH product and the agent being tested, may be measured. Alternatively, the activator or deactivator may work by serving as agonist or antagonist, respectively, of the CLH receptor and their effect may be determined in connection with the receptor.

Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the CLH product is described in detail by Geysen in PCT Application WO 84/03564, published on Sep. 13, 1984. In summary, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with the full CLH product or with fragments of CLH product and washed. Bound CLH product is then detected by methods well known in the art. Substantially purified CLH product can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

Antibodies to the CLH product, as described in Example IV below, may also be used in screening assays according to methods well known in the art. For example, a *"sandwich"* assay may be performed, in which an anti-CLH antibody is affixed to a solid surface such as a microtiter plate and CLH product is added. Such an assay can be used to capture compounds which bind to the CLH product. Alternatively, such an assay may be used to measure the ability of compounds to influence with the binding of CLH product to the CLH receptor [I and then select those compounds which effect the binding.

Example IV. Anti-CLH antibodies

A. Synthesis

In still another aspect of the invention, the purified CLH product is used to produce anti-CLH antibodies which have diagnostic and therapeutic uses related to the activity, distribution, and expression of the CLH product, in particular

therapeutic applications mentioned in connection with the pharmaceutical composition aspect of the invention.

Antibodies to CLH product may be generated by methods well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, humanized, single chain, Fab fragments and fragments produced by an Fab expression library. Antibodies, i.e., those which inhibit dimer formation, are especially preferred for therapeutic use.

A fragment CLH product for antibody induction does not require biological activity but have to feature immunological activity; however, the protein fragment or oligopeptide must be antigenic. Peptides used to induce specific antibodies may have an amino acid sequence consisting of at least five amino acids, preferably at least 10 amino acids of the sequences specified in SEQ ID NO: 12 to SEQ ID No. 22. Preferably they should mimic a portion of the amino acid sequence of the natural protein and may contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of CLH protein amino acids may be fused with those of another protein such as keyhole limpet hemocyanin and antibody produced against the chimeric molecule. Procedures well known in the art can be used for the production of antibodies to CLH product.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, etc may be immunized by injection with CLH product or any portion, fragment or oligopeptide which retains immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include but are not limited to Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are potentially useful human adjuvants.

Monoclonal antibodies to CLH protein may be prepared using any technique which provides for the production of antibody molecules by continuous

cell lines in culture. These include but are not limited to the hybridoma technique originally described by Koehler and Milstein (*Nature* **256**:495-497, (1975)), the human B-cell hybridoma technique (Kosbor *et al., Immunol. Today* **4**:72, (1983); Cote *et al., Proc. Natl. Acad. Sci.* **80**:2026-2030, (1983)) and the EBV-hybridoma technique (Cole, *et al., Mol. Cell Biol.* 62:109-120, (1984)).

Techniques developed for the production of *"chimeric antibodies"*, the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can also be used (Morrison *et al., Proc. Natl. Acad. Sci.* **81**:6851-6855, (1984); Neuberger *et al., Nature* **312**:604-608, (1984); Takeda *et al., Nature* **314**:452-454, (1985)). Alternatively, techniques described for the production of single chain antibodies (U.S. Pat. No. 4,946,778) can be adapted to produce single-chain antibodies specific for the CLH protein.

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening recombinant immunoglobulin libraries or panels of highly specific binding reagents as disclosed in Orlandi *et al.* (*Proc. Natl. Acad. Sci.* **86**:3833-3837, 1989)), and Winter G and Milstein C., (*Nature* **349**:293-299, (1991)).

Antibody fragments which contain specific binding sites for CLH protein may also be generated. For example, such fragments include, but are not limited to, the $F(ab')_2$ fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse W.D. *et al., Science* 256:1275-1281, (1989)).

B. Diagnostic applications of antibodies

A variety of protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established

specificities are well known in the art. Such immunoassays typically involve the formation of complexes between CLH product and its specific antibody and the measurement of complex formation. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two noninterfering epitopes on a specific CLH product is preferred, but a competitive binding assay may also be employed. These assays are described in Maddox D.E., *et al.*, (*J. Exp. Med.* **158**:1211, (1983)).

Antibodies which specifically bind CLH product are useful for the diagnosis of conditions or diseases characterized by over or under expression of CLH. Alternatively, such antibodies may be used in assays to monitor patients being treated with CLH product, its activators, or its deactivators. Diagnostic assays for CLH protein include methods utilizing the antibody and a label to detect CLH product in human body fluids or extracts of cells or tissues. The products and antibodies of the present invention may be used with or without modification. Frequently, the proteins and antibodies will be labeled by joining them, either covalently or noncovalently, with a reporter molecule. A wide variety of reporter molecules are known in the art.

A variety of protocols for measuring CLH product, using either polyclonal or monoclonal antibodies specific for the respective protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescent activated cell sorting (FACS). As noted above, a two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on CLH product is preferred, but a competitive binding assay may be employed. These assays are described, among other places, in Maddox, *et al. (supra)*. Such protocols provide a basis for diagnosing altered or abnormal levels of CLH product expression. Normal or standard values for CLH product expression are established by combining body or cell extracts taken from normal subjects, preferably human, with antibody to CLH product under conditions suitable for complex formation which are well known in the art. The amount of standard complex formation may be quantified

by various methods, preferably by photometric methods. Then, standard values obtained from normal samples may be compared with values obtained from samples from subjects potentially affected by disease. Deviation between standard and subject values establishes the presence of disease state.

The antibody assays are useful to determine the level of CLH present in a body fluid sample, in order to determine whether it is being overexpressed or underexpressed in the tissue, or as an indication of how CLH levels are responding to drug treatment.

Another alternative is to determine the presence and/or level of naturally occurring anti-CLH antibodies in a sample, such as blood or serum. Many times diseases are identified by detecting the presence or level of antibodies against a specific product. For the detection of such naturally occurring anti-CLH antibodies, the sample may be contacted with the product of the invention, for example as depicted in any one of SEQ ID NO: 5 to SEQ ID NO: 8, or with an antigenic fragment thereof, and the presence or level of antibody-antigen complexes may be determined by methods well known in the art.

C. Therapeutic uses of antibodies

In addition to their diagnostic use the antibodies may have a therapeutical utility in blocking or decreasing the activity of the CLH product in pathological conditions where beneficial effect can be achieved by such a decrease.

The antibody employed is preferably a humanized monoclonal antibody, or a human Mab produced by known globulin-gene library methods. The antibody is administered typically as a sterile solution by IV injection, although other parenteral routes may be suitable. Typically, the antibody is administered in an amount between about 1-15 mg/kg body weight of the subject. Treatment is continued, e.g., with dosing every 1-7 days, until a therapeutic improvement is seen.

Although the invention has been described with reference to specific methods and embodiments, it is appreciated that various modifications and changes may be made without departing from the invention.

Example V: Experimental Procedures

A. RNA Purification and cDNA Synthesis

Total RNA was extracted from different human tissues using Tri-Reagent System (Molecular Research Center, Inc., Cincinnati, OH).

Poly (A) RNA was isolated from total RNA using Dynabeads mRNA Direct Kit (Dynal, Oslo, Norway).

Synthesis of first-strand cDNA was carried out using Oligo$(dT)^{15}$ (Promega, Medison, WI), Superscript II or ThermoScript RNase H$^-$ Reverse transcriptase (Gibco/BRL, Gaithersburg, MD), Rnasin (Promega, Medison, WI) and dNTP's (Gibco/BRL, Gaithersburg, MD).

B. RACE analysis of 5' and 3' ends of LM

5' and 3' RACE (rapid amplification of cDNA ends) analysis was performed on poly A RNA from human placenta tissue using the Marathon cDNA Amplification Kit (Clontech). Adaptor-ligated double-stranded cDNA libraries were prepared essentially as suggested by the manufacturer. Superscript II Reverse Transcriptase (Gibco/BRL, Gaithersburg, MD) was used for the first strand synthesis. First round PCR was performed on these libraries for 30 cycles, using the Expand Long Template PCR System (Boerhringer-Mannheim, Germany). A nested PCR approach was used to isolate 5' and 3' RACE products.

C. Polymerase Chain Reaction (PCR):

PCR was performed using either Taq DNA polymerase or Expand Long Template PCR system (Roche) pretreated with Taq Start Antibody (Hot

start). As a template we used cDNA from different tissues. The PCR reaction on PTC-225 (MJ Research, Inc.). PCR products were analyzed on an automated DNA sequencer ABI Prizem 310 Genetic Analyzer (Perkin Elmer).

D. Northern blot

Random primer DNA labeling was performed using ^{32}P and MIT (Biological Industries Co., Beit Haemek LTD). Chordin-LM probes used were a product of the sense primer:

5'- GAAAGCCTGTGTGCATGGCGG-3'

and the anti-sense primer:

5'-AGCTCATATCTGCAACTGTTAGG-3'.

The membranes used were from Human Muscle Multiple Tissue Northern Blot (MTN™, Clontech).

E. Expressiont of GST-LM

The PCR product was cloned into plasmid PGEX-6p (Pharmacia Biotech) and expressed in E.coli DH5-alfa as a fusion protein with GST. Expression, purification and detection of the fusion protein GST- LM was performed following the manufacturer's instructions.

F. Preparation of Antibodies

The anti- LM was prepared by immunizing rabbits with the purified fusion GST-LM.

G. Immunohistochemistry

Immunohistochemical staining was performed using Histostain plus Kit (Zymed Laboratories Inc.). Different human micron sections were prepared using a R. Gung microtome and fixed on superfrost plus slides with 2% Tespa. Deparaffinization was performed in xylene for 10 min.

Dehydration was performed three times in absolute ethanol and once 95% ethanol. The slides were washed in DDW and then icubated with 3% H_2O_2 for 5 min. Subsequently, the slide were washed in DDW and twice in 0.05M TrisHCl pH 7.6 (Optimax wash Buffer, BioGenex). The rest of the procedure was performed following the manufacturer's instructions.

H. Expression Plasmids:

The variants were cloned into pCDNA3 mammalian expression vector (Invitrogen).

I. Transfection experiments:

Chordin-LM was transiently expressed in COS-7 cell line (ATCC). The transfection of the expression vector into COS-7 was done by the FuGENE™ 6 method according to the manufacturer's instructions (Boeringer Mannheim).

J. Western blot

Protein samples were separated by SDS-PAGE, electrophoretically transferred to a nitrocellulose membrane (Pharmacia Biotech), and subjected to immunodetection using the immunized sera as a primary antibody and peroxidase-conjugated Gout Anti Rabbit IgG (Jackson Immunoresearch Laboratories, Inc.). Proteins were visualized with enhanced chemiluminescence system (Pierce).

EXAMPLE VI: Expression pattern of CLH

To begin characterization of CLH RNA expression, Northern blot analysis was performed. The membranes used were from Human Muscle Multiple Tissue Northern Blot (MTN™, Clontech). CLH mRNA of 2.3kb was detected at

significantly high levels in uterus, and also in colon, bladder, heart, stomach and prostate as shown in Fig. 22.

Expression of CLH mRNA was also found in different human cDNA tissues, such as: testis, placenta, brain, bone marrow, ovary, fetal lung, fetal brain. (data not shown).

EXAMPLE VII: Generation of CLH specific antibodies:

In order to generate antibodies against CLH, DNA fragment containing Chordin-like variant 1 was PCR-amplified and cloned into pGEX-6p vector (Pharmacia Biotech). Using the glutathione S- transferase (GST) gene fusion system, CLH fused to GST was expressed, purified and detected on SDS – PAGE. Large scale of CLH fused protein was prepared to immunized rabbits. Sera before the rabbits immunization was collected (reffered as the pre-immuned Ab's) and also following serial rabbit immunization with the purified fusion GST-LM (reffered as the anti-LM Ab's). The antibodies thus produced were used for immunohistochemical studies.

EXAMPLE VIII: Expression and secretion of CLH in mammalian cell line

As mentioned before, CLH SEQ ID No. 16 contains predicted signal peptide. In order to validate secretion of the protein, DNA fragment containing CLH SEQ ID No. 5 was PCR-amplified and cloned into pCDNA3 mammalian expression vector (Invitrogen).

COS-7 cells were transfected with pCDNA3 carrying CLH gene or with pCDNA3 alone. After incubating for 48 hr and 72 hr, mediums from the transfected cells were collected and protein extraction from the cells was performed. Protein samples from both the cell lysate and the medium were separated by SDS-PAGE, electrophoretically transferred to a nitrocellulose membrane and subjected to immunodetection using the anti- LM Ab's of example VII . The expression and the secretion of the CLH SEQ ID No. 13

variant 1 molecule is presented in figure 23.

As shown in Fig 22 , lanes A3, B3, C3, COS7 untransfected cells (referred to as *Mock)* do not express CLH endogenously. CLH was over expressed only in the cells transfected with pCDNA3 carrying CLH gene Fig 23 , lane C1 and not in the cells transfected with pCDNA3 Fig 23 , lane C2. Moreover, high levels of secreted protein were detected in the medium of CLH transfected cells following 48hr and 72 hr (Fig22 Lanes A1 and lane B1 respectively), and not in the cells transfected with pCDNA3 (Fig22 Lanes A2 and lane B2 respectively).

EXAMPLE IX: Immunohistochemical localization of CLH protein in different human tissues:

Immunohistochemical staining was performed on different human micron sections using the anti-LM antibodies (Fig 24 right column letters with prime) indicated compared to the pre-immune rabbit's serum (Fig 24, left columns, indicated in normal letters). CLH was found to be expressed in different epithelial tissues (Fig.24 a', b', c', d', e', f', g') and localized mainly in the secreting cells.

Expression of CLH was detected in ductal epithelium of the breast. Breast carcinoma was positively stained both in the regions of ductal carcinoma (Fig. 24 a') in situ (DCIS) and of invasive ductal carcinoma (Fig. 24b). Secreting cells in benign prostatic hyperplasia (BPH) and prostate carcinoma sections were also positively stained Fig. 24, c', d', respectively.

CLH was localized to the transitional epithelium in the bladder (Fig 24 e'). The internal female genitalia (fallopian tube, endocervical glands and the uterus) which evolved from the same embryonic precursor - the mullerian duct, showed positive staining (Fig.24, e'). Expression of CLH was localized in the lining epithelium of the fallopian tube (Fig 24, f'), in the endocervical glands (Fig 24, g') and in the normal and endometrial carcinoma of the uterus (Fig. 24, h' and i', respectively). However, in the region of the mucinous metaplasia in the endometrial carcinoma, negative staining of CLH was observed

(Fig 24, j').

CLH was localized not only in epithelial tissues as mentioned above, but also in osteoblasts in the fetal bone of thigh (Fig 24k').

Positive staining of CLH was also detected in activated astrocyte (referred as Gemistocyte , Fig 24l') in Glioblastoma Multiforme-GBM (brain tumor) but not in oligodendroglia (Fig 24l negative staining).

EXAMPLE X: CLH protein expression - Western Blot Analysis

CLH was detected in different tissues by Western blot analysis using anti-LM Ab's. As shown in Fig. 25, CLH is expressed in the brain and bone tumor (Fig 25A and 2B respectively). CLH in the transfection medium (experiment described in details previously), served as a positive control (Fig 25 refered as positive control). Multiple bands in the Western blot analysis may reflect either alternative splicing products of a single gene or post translational modifications (PTM) of CLH.

CLAIMS:

1. An isolated nucleic acid sequence selected from the group consisting of:

(i) the nucleic acid sequence depicted in any one of SEQ ID NO: 1 to SEQ ID NO: 11;

(ii) nucleic acid sequences having at least 70% identity with the sequence of (i); and

(iii) fragments of (i) or (ii) of at least 20 b.p.

2. A nucleic acid sequence according to Claim 1(ii) wherein the nucleic acid sequences have at least 80% identity with the sequence of Claim 1(i).

3. A nucleic acid sequence according to Claim 2, wherein the nucleic acid sequences have at least 90% identity.

4. A nucleic acid sequence according to Claim 3, wherein the nucleic acid sequences have at least 95% identity.

5. An isolated nucleic acid sequence complementary to the nucleic acid sequence of Claim 1.

6. An amino acid sequence selected from the group consisting of:

(i) an amino acid sequence coded by the isolated nucleic acid sequence of Claim 1;

(ii) fragments of the amino acid squence of (i) having at least 10 amino acids;

(iii) analogues of the amino acid sequences of (i) or (ii) in which one or more amino acids has been added, deleted, replaced or chemically modified without substantially altering the biological activity of the parent amino acid sequence.

7. An amino acid sequence according to Claim 6, as depicted in any one of SEQ ID NO: 12 to SEQ ID NO: 22.

8. An isolated nucleic acid sequence coding for the amino acid sequence of Claim 6 or 7.

9. A purified antibody which binds specifically to the amino acid sequence of Claim 6 or 7.

10. An expression vector comprising the nucleic acid sequences of Claim 1 or 8 and control elements for the expression of the nucleic acid sequence in a suitable host.

11. An expression vector comprising the nucleic acid sequence of Claim 5, and control elements for the expression of the nucleic acid sequence in a suitable host.

12. A host cell transfected by the expression vector of Claim 10 or 11.

13. A pharmaceutical composition comprising a pharmaceutically acceptable carrier and as an active ingredient an agent selected from the group consisting of:

(i) the expression vector of Claim 10; and

(ii) the amino acid sequence of Claim 6 or 7.

14. A pharmaceutical composition according to Claim 13, for treatment of diseases which can be ameliorated, cured or prevented by raising the level of a Chordin-Like-Homolog (CLH).

15. A pharmaceutical composition comprising a pharmaceutically acceptable carrier and as an active ingredient an agent selected from the group consisting of:

(i) the nucleic acid sequence of Claim 5;

(ii) the expression vector of Claim 11; and

(iii) the purified antibody of Claim 9.

16. A pharmaceutical composition according to Claim 15, for treatment of diseases which can be ameliorated or cured by decreasing the level of the CLH product.

17. A pharmaceutical composition according to Claim 13 or 15, for the treatment of a disease selected from: diseases manifested in non-normal bone formation and non-normal bone modeling; bone injuries; diseases involved with the female reproductive tract; diseases of disorders involved with abnormal sexual differentiation; recurrent miscarriages, tumors of the uterus, breast or prostate; diseases involving sexual hormone abnormalities; cardiovascular disorders;

neuronal diseases of the CNS; neurode-generative diseases and diseases involving non-normal developments of neurons.

18. A method for detecting an CLH nucleic acid sequence in a biological sample, comprising the steps of:

(a) hybridizing to nucleic acid material of said biological sample a nucleic acid sequence of Claim 1 or 5; and

(b) detecting said hybridization complex;

wherein the presence of said hybridization complex correlates with the presence of an CLH nucleic acid sequence in the said biological sample.

19. A method according to Claim 18, wherein the nucleic acid material of said biological sample are mRNA transcripts.

20. A method according to Claim 18, where the nucleic acid sequence is present in a nucleic acid chip.

21. A method for identifying candidate compounds capable of binding to the CLH product and modulating its activity the method comprising:

(i) providing a protein or polypeptide comprising an amino acid sequence substantially as depicted in any one of SEQ ID NO: 12 to SEQ ID NO: 22, or a fragment of such a sequence;

(ii) comparing the physiological effect of the CLH product in the absence and presence of said candidate compound and selecting those compounds which show a significant effect on said physiological activity.

22. A method according to Claim 21, wherein the compound is an activator and the measured effect is increase in the physiological activity.

23. A method according to Claim 21, wherein the compound is an deactivator and the effect is decrease in the physiological activity.

24. An activator of the amino acid sequence of Claim 6 or 7.

25. An deactivator of the amino acid sequence of Claims 6 or 7.

26. A method for detecting CLH-product in a biological sample, comprising the steps of:

(a) contacting with said biological sample the antibody of Claim 9, thereby forming an antibody-antigen complex; and

(b) detecting said antibody-antigen complex

wherein the presence of said antibody-antigen complex correlates with the presence of CLH product in said biological sample.

27. A method for detecting anti-CLH antibodies in a biological sample comprising the steps of:

(a) contacting with said biological sample the antibody of Claim 6 or 7, thereby forming an antibody-antigen complex; and

(b) detecting said antibody-antigen complex

wherein the presence of said antibody-antigen complex correlates with the presence of anti-CLH antibody in said biological sample.

```
 26 TGCCTGCGCTGTACCTGCTCAGAGGGCGCCCATGTGAGTTGTTACCGCCT 75
    |||:::   |||||||||:::::   :::   |||   |||     ::
711 CysPheThrCysThrCysGlnLysLysThr...ValIleCysAspProVa 726

 76 CCACTGTCCGCCTGTCCACTGCCCCCAGCCTGTGACGGAGCCACAGCAAT 125
    :   ||||||   :::   |||   :::   |||         :::||||
726 lMetCysProThrLeuSerCysThrHisThrValGlnProGluAspGlnC 743

126 GCTGTCCCAAGTGTGTGGAACCTCACACTCCCTCTGGACTCCGGGCC... 172
     ||||||||   |||   |||              :::        |||
743 ysCysProIleCysGluGluLysLysGluSerLysGluThrAlaAlaVal 759

173 ..............CCACCAAAGTCCTGCCAGCACAACGGGACCATGTA 207
                |||:::::::|||      :::|||   :::
760 GluLysValGluGluAsnProGluGlyCysTyrPheGluGlyAspGlnLy 776

208 CCAACACGGAGAGATCTTCAGTGCCCATGAGCTGTTCCCCTCCCGCCTGC 257
       :::|||:::          :::   |||   :::   |||
776 sMetHisAlaProGlyThrThrTrpHisProPheValProProPheGlyT 793

258 CCAACCAGTGTGTCCTCTGCAGCTGC......ACAGAGGGCCAGATCTAC 301
      :::|||::::::|||:::|||      :::   |||:::::::::
793 yrIleLysCysAlaValCysThrCysLysGlySerThrGlyGluValHis 809
```

FIG. 1

```
302 TGCGGCCTCACAACCTGCCCCGAACCAGGCTGCCCAGCACCCCTCCCGCT 351
      |||      :::||||||||||          |||      |||:::
810 CysGluLysValThrCysProProLeuThrCysSerArgProIleArgAr 826

352 G...CCAGACTCCTGCTGCCAAGCCTGCAAAGATGAGGCAAGTGAGCAAT 398
       |||::::::||||||:::   |||      |||   :::
826 gAsnProSerAspCysCysLysGluCysProProGluGluThrProProL 843

399 CGGATGAAGAGGACAGTGTGCAGTCGCTCCATGGGGTGAGACATCCTCAG 448
                 ::::::|||:::   :::|||:::
843 euGluAspGluGluMetMetGlnAla..................... 851

449 GATCCATGTTCCAGTGATGCTGGGAGAAAGAGAGGCCCGGGCACCCCAGC 498
                              ||||||
852 ..............................AspGlyThr..... 854

499 CCCCACTGGCCTCAGCGCCCCTCTGAGCTTCATCCCTCGCCACTTCAGAC 548

854 ..................................... 854

549 CCAAGGGAGCAGGCAGCACAACTGTCAAGATCGTCCTGAAGGAGAAACAT 598

854 ..................................... 854
```

FIG. 1 (CONT.[1])

```
599 ANGAAAGCCTGTGTGCATGGCGGGAAGACGTACTCCCACGGGGAGGTGTG 648
     :::   |||      |||   :::   |||:::::::::::|||   ||
855 ...ArgLeuCysLysPheGlyLysAsnTyrTyrGlnAsnSerGluHisTr 870

649 GCACCCGGCCTTCCGTGCCTTCGGCCCTTGCCCATGCATCCTATGCACCT 698
    |||||||:::           |||          ||||||  |||   |
870 pHisProSerValProLeuValGlyGluMetLysCysIleThrCysTrpC 887

699 GTGAGGATGGCCGCCAGGACTGCCAGCGTGTGACCTGTCCCACGAAGTAC 748
        ||:::   |||         |||||||||      ||||||
887 ysAspHisGlyValThrLysCysGlnArgLysGlnCysProLeu...Leu 902

749 CCCTGCCGTCACCCCGAGAAAGTGGCTGGGAAGTGCTGCAAGATTTGCCC 798
       ||||||:::|||   ::::::   ||||||||||||      |||
903 SerCysArgAsnProIleArgThrGluGlyLysCysCysProGluCysIl 919

799 AGAGGAC 805
    |||||||
919 eGluAsp 921
```

FIG. 1 (CONT.[2])

```
863 TCGCTCCATGGGGTGAGACATCCTCAGGATCCATGTTCCAGTGATGCTGG 912
    |||   |||      ::::::          |||      :::   :::||
  3 SerTyrHisArgSerHisTyrAspProProProSerArgGlnAlaGlyGl 19

913 GAGAAAGAGAGGCCCGGGCACCCCAGCCCCCACTGGCCTCAGCGCCCCTC 962
          |   :::|||   ||||||:::   :::      |||
 19 yLeuSerArgPheProGlyAlaArgSerHisArgGlyAlaLeu..... 33

963 TGAGCTTCATCCCTCGCCACTTCAGACCCAAGGGAGCAGGCAGCACAACT 1012
                :::      :::   |||:::::::|||
  34 ....................MetAspSerGlnGlnAlaSerGlyThrIle 43

1013 GTCAAGATCGTCCTG......AAGGAGAAACATANGAAAGCCTGTGTGCA 1056
     |||:::|||||||:::      |||:::||||||   ::::::||||||
  44 ValGlnIleValIleAsnAsnLysHisLysHisGlyGlnValCysValSe 60

1057 TGGCGGGAAGACGTACTCCCACGGGGAGGTGTGGCACCCGGCCTTCCGTG 1106
     :::|||||||||||||||||||||||   ||||||||   :::||||
  60 rAsnGlyLysThrTyrSerHisGlyGluSerTrpHisProAsnLeuArgA 77

1107 CCTTCGGCCCTTGCCCATGCATCCTATGCACCTGTGAGGATGGCCGCCAG 1156
     ||||||||          |||:::|||||||||||||:::      :::|||
  77 laPheGlyIleValGluCysValLeuCysThrCysAsnValThrLysGln 93
```

FIG.2A

```
1157 GACTGCCAGCGTGTGACCTGTCCCACGAAGTACCCCTGCCGTCACCCCGA 1206
     :::|||:::::::::   ||||||:::::|||||||||:::::|||::
  94 GluCysLysLysIleHisCysProAsnArgTyrProCysLysTyrProGl 110

1207 GAAAGTGGCTGGGAAGTGCTGCAAGATTTGCCCAGAGGACAAAGCAGAC. 1255
       :|||:::   ||||||||||||||::||||||      ||||||
  110 nLysIleAspGlyLysCysCysLysValCysProGlyLysLysAlaLysG 127

1256 .........CCTGGCCACAGT...GAGATCAGTTCTACCAGGTGTCCCAAG 1294
         ||||||:::|||   :::   ::::::      |||   :::
  127 luGluLeuProGlyGlnSerPheAspAsnLysGlyTyrPheCysGlyGlu 143

1295 GCACCGGGCCGGGTCCTCGTCCACACATCGGTA...TCCCCAAGCCCAGA 1341
           :::   |||:::  ||||||          :::   ::
  144 Glu......ThrMetProValTyrGluSerValPheMetGluAspGlyGl 158

1342 CAACCTGCGTCGCTTTGCCCTGGAACACGAGGCCTCGGACTTGGTGGAGA 1391
     ::::   |||::::::|||||||||   |||                |||||||:
  158 uThrThrArgLysIleAlaLeuGluThrGluArgProProGlnValGluV 175

1392 TCTACCTCTGGAAGCTGGTAAANNNNNNNNNNNNNNNNNNNNNNNNNNN 1441
                      :::::::::|||   :::
  175 alHisValTrpThrIle..............................  180
```

FIG.2A (CONT.[1])

```
1442 NNNNNNNNNNNNNNNNNNNNNNNNNNNCAGAATTTCCACTTGACTCAGAT 1491
                  |||:::||||||          ||
 181 ............ArgLysGlyIleLeuGlnHisPheHis.........Il 190

1492 CAAGAAAGTCAGGAAGCAAGACTTCCAGAAAGAGGCACAGCACTTCCGAC 1541
     |:::|||:::   |||:::   |||:::   |||      ||||||:::|
 190 eGluLysIleSerLysArgMetPheGlu...GluLeuProHisPheLysL 206

1542 TGCTCGCTGGCCCCCACGAAGGTCACTGGAACGTCTTCCTAGCCCAGACC 1591
           ||::::::             ::::::|||:::::::|||
 206 euValThrArgThrThrLeuSerGlnTrpLysIlePheThrGluGlyGlu 222

1592 CTGGAGCTGAAGGTCACGGCCAGTCCAGACAAAGTGACCAAGACATAACA 1641
        :::::::::      :::|||                      +++
 223 AlaGlnIleSerGlnMetCysSer.......................... 230

1642 AAGACCTAACAGTTGCAGATATGAGCTGTATAATTGTTGTTATTATATAT 1691
         :::   ||||||   ||||||   :::             :::
 231 ....SerArgValCysArgThrGluLeuGluAspLeuValLysValLeuT 246

1692 TAATAAATAAGAAGTTGCATAACCAT 1717
      ::::::::::::      :::|||
 246 yrLeuGluArgSerGluLysGlyHis 254
```

FIG.2A (CONT.[2])

```
100 TCTTCCCACCTTAGACCTCCCTTCCTGCCCTCCTTTCCTGCCCACCGCTG 149
     ::::::||||||      |||      |||:::   |||:::   |||
429 ThrAlaHisLeuLeuGlyPro......ProGlyThrProGlyProArg.. 442

150 CTTCCTGGCCCTTCTCCGACCCCGCTCTAGCAGCAGACCTCCTGGGGTCA 199
         ||||||:::          ::::|||:::      ||||||:
443 .......ArgLeuLeuLysGlyPheTyrGlySerGluAlaGlnGlyValV 457

200 TGTGGGTTGATCTGTGGCCCCTGTGNCTCCGTGTCCTTTTCGTCTCCCGT 249
          ::      ||||||   |||      ||||||   |||
457 al...LysAspLeuGluProGluLeuLeuArgHisLeu..... 468

250 CCTCCCGACTCCGCTCCCGGACCAGCGGCCTGACCCTGGGGAAAGGATGG 299
          |||   |||   |||:::+++                  ||
469 ............AlaLysGlyMetAlaSerLeuLeuIleThrThrLysGl 481

300 TTCC...CGAGGTGAG..................................... 312
                    ||||   |||||||||
481 ySerProArgGlyGluLeuArgGlyGlnValHisIleAlaAsnGlnCysG 498
```

FIG. 2B

```
313 .....GGTCCTCTCCTCCTT.................GCTGGGACTCGCGCT 342
      |||   |||   |||                      |||:::||||||
498 luValGlyGlyLeuArgLeuGluAlaAlaGlyAlaGluGlyValArgAla 514

342 ..................................................... 342

515 LeuGlyAlaProAspProAlaSerAlaAlaProProValValProGlyLe 531

343 ....GCTCTGGTTCCC......CCTGGACTCCCACGCTCGAGCCCGCCCA 382
        ||||||:::|||      ||||||   ||| :::|||   |||
531 uProAlaLeuAlaProAlaLysProGlyGlyPro.GlyArgProArgAsp 547

383 GACATGTTCTGCCTTTTCCATGGGAAGAGATACTCCCCCGGCGAGAGCTG 432
       |||:::|||:::|||::::::          |||      ||
548 ProAsnThrCysPhePheGluGlyGlnGlnArgProHisGlyAlaArgTr 564

433 GCACCCCTACTTGGAGCCACAAGGCCTGATGTACTGCCTGCGCTGTACCT 482
    |   |||      :::|||                |||     |||||||
564 pAlaProAsnTyrAspPro.............LeuCysSerLeuCysThrC 577
```

FIG. 2B (CONT.[1])

```
483 GCTCAGAGGGCGCCCATGTGAGTTGTTACCGCCTCCACTGTCCGCCTGTC 532
    ||::::::   :::   |||   |||      :::   |||||||||
577 ysGlnArgArgThr...ValIleCysAspProValValCysProProPro 592

533 CACTGCCCCCAGCCTGTGACGGAGCCACAGCAATGCTGTCCCAAGTGTGT 582
    ||||||::::||||||      |||:::|||||||||||||   |||
593 SerCysProHisProValGlnAlaProAspGlnCysCysProValCysPr 609

583 GGAA.....CCTCACACTCCCTCTGGACTCCGGG 611
    |||          |||      |||   :::   |||
609 oGluLysGlnAspValArgAspLeuProGlyLeuProArgSer...ArgA 625

612 CCCCACCAAAGTCCTGCCAGCACAACGGGACCATGTACCAACACGGAGAG 661
      |||   :::::::|||      :::|||            ::::::
625 spProGlyGluGlyCysTyrPheAspGlyAspArgSerTrpArgAlaAla 641

662 ATCTTCAGTGCCCATGAGCTGTTCCCCTCCCGCCTGCCCAACCAGTGTGT 711
        |||   :::   |||                ::::|||::
642 GlyThrArgTrpHisProValValProProPheGlyLeuIleLysCysAl 658

712 CCTCTGCAGCTGCACA......GAGGGCCAGATCTACTGCGGCCTCACAA 755
    ::::|||:::|||            |||::::::::::|||      :::
658 aValCysThrCysLysGlyGlyThrGlyGluValHisCysGluLysValG 675
```

FIG. 2B (CONT.[2])

```
756 CCTGCCCCGAACCAGGCTGCCCAGCACCCCTCCCGCTG...CCAGACTCC 802
     ||||||:::   :::|||      |||:::   :::   |||   :::
675 lnCysProArgLeuAlaCysAlaGlnProValArgValAsnProThrAsp 691

803 TGCTGCCAAGCCTGCAAAGATGAGGCAAGTGAGCAATCGGATGAAGAGGA 852
                    ||||||:::   |||
692 CysCysLysGlnCys..............................  696

853 CAGTGTGCAGTCGCTCCATGGGGTGAGACATCCTCAG......GATCCAT 896
          :::   :::      |||||||||      ||||||
697 ..........ProValGlySerGlyAlaHisProGlnLeuGlyAspProM 710

897 GTTCCAGTGATGCTGGGAGAAAGAGAGGCCCGGGCACCCCAGCCCCCACT 946
               ::::::|||                |||||
710 etGlnAlaAsp..............GlyPro.................. 715

947 GGCCTCAGCGCCCCTCTGAGCTTCATCCCTCGCCACTTCAGACCCAAGGG 996

715 ..........:..................................... 715

997 AGCAGGCAGCACAACTGTCAAGATCGTCCTGAAGGAGAAACATANGAAAG 1046
                          ::::
716 .....................................ArgG 717
```

FIG. 2B (CONT.[3])

```
1047 CCTGTGTGCATGGCGGGAAGACGTACTCCCACGGGGAGGTGTGGCACCCG 1096
     ::|||      :::|||:::   :::   :::::::::   |||||||||
 717 lyCysArgPheAlaGlyGlnTrpPheProGluSerGlnSerTrpHisPro 733

1097 GCCTTCCGTGCCTTCGGCCCTTGCCCATGCATCCTATGCACCTGTGAGGA 1146
       :::          ||||||          ||||||  |||   |||
 734 SerValProProPheGlyGluMetSerCysIleThrCysArgCysGlyAl 750

1147 TGGCCGCCAGGACTGCCAGCGTGTGACCTGTCCCACGAAGTACCCCTGCC 1196
      |||         |||:::|||      |||                |||
 750 aGlyValProHisCysGluArgAspAspCysSerLeuProLeuSerCysG 767

1197 GTCACCCCGAGAAAGTGGCTGGGAAGTGCTGCAAGATTTGC..... 1237
           ::::::      ::::::|||||||:::   |||
767 lySerGlyLysGlu......SerArgCysCysSerArgCysThrAlaHis 781

1238 .............CCAGAGGACAAAGCAGACCCT 1258
                |||||   ::::::||||||
 782 ArgArgProAlaProGluThrArgThrAspPro 792
```

FIG. 2B (CONT.[4])

```
  3 CTTCCCCCTTTCTTTGATCGCCTCTCC.............CTTCTGCTGGA 40
    |||||||||      ::::::   :::              |||   |||||
540 LeuProProValArgSerGlnAlaAlaGlyHisAlaTrpLeuSerLeuAs 556

 41 CCTTCCTTCGTCTCTCCATCTCTCCCTCCTT...................T 72
    |             ||||||   :::::::|||                   |
556 pThrHisCysHisLeuHisTyrGluValLeuLeuAlaGlyLeuGlyGlyS 573

 73 CCCCGCGTTCTCTTTCCACCTTTCTCTTCTTCCCACCTTAGACCTCCCTT 122
       ||   ::::::      |||                    +++
573 erGluGlnGlyThrValThr.............................. 579

123 CCTGCCCTCCTTTCCTGCCCACCGCTGCTTCCTGGCCCTTCTCCGACCCC 172
     |||   |||      ||||||   :::||||||||||::: |||
580 ...AlaHisLeuLeuGlyProProGlyMetProGlyProGln.ArgLeuL 595

173 GCTCT.........AGCAGCAGACCTCCTGGGGTCATGTGGGTTGATCTG 213
     :::          :::|||:::      ||||||:::      ||||||
595 euLysGlyPheTyrGlySerGluAlaGlnGlyValVal...LysAspLeu 610

214 TGGCCCCTGTGNCTCCGT.............................GT 233
     |||:::   ||||||                                ::
611 GluProValLeuLeuArgHisLeuAlaGlnGlyThrAlaSerLeuLeuIl 627
```

Fig. 2C

```
234 CCTTTTCGTCTCCCGTCCTCCCGACTCCGCTCCCGGACCA..... 273
          :          |||   |||      :::      |||
627 eThrThrLysSerSerProArgGlyGluLeuArgGlyGlnValHisIleA 644

274 .....................................GCGGCC 279
                                  :::
644 laSerGlnCysGluAlaGlyGlyLeuArgLeuAlaSerGluGlyValGln 660

280 TGACCCTGGGGAAAGGATGGTTCCCGAGGTGAGGGTCCTCTCCTCCTTGC 329
    +++|||   :::   :::|||:::   :::   :::|||:::|||   ||
661 MetProLeuAlaProAsnGlyGluAlaAlaThrSerProMetLeuProAl 677

330 TGGGACT...CGCGCTGCTCTGGTTCCCCCT......GGACTCCCACGCT 370
    ||||          |||   ||||||          |||   ||| :::
677 aGlyProGlyProGluAlaProValProAlaLysHisGlySerPro.Gly 693

371 CGAGCCCGCCCAGACATGTTCTGCCTTTTCCATGGGAAGAGATACTCCCC 420
       |||   |||            |||:::|||:::|||:::::
694 ArgProArgAspProAsnThrCysPhePheGluGlyGlnGlnArgProHi 710

421 CGGCGAGAGCTGGCACCCCTACTTGGAGCCACAAGGCCTGATGTACTGCC 470
     |||      |||   |||      :::|||                |||
710 sGlyAlaArgTrpAlaProAsnTyrAspPro.............LeuCysS 723
```

FIG. 2C (CONT.[1])

```
471 TGCGCTGTACCTGCTCAGAGGGCGCCCATGTGAGTTGTTACCGCCTCCAC 520
       |||   |||::::::   :::   |||   |||      :::
723 erLeuCysIleCysGlnArgArgThr...ValIleCysAspProValVal 738

521 TGTCCGCCTGTCCACTGCCCCCAGCCTGTGACGGAGCCACAGCAATGCTG 570
    |||||||||     ||||||:::||||||          :::|||||||||
739 CysProProProSerCysProHisProValGlnAlaLeuAspGlnCysCy 755

571 TCCCAAGTGTGTGGAA.....CCTCACACTCCCT 599
    ||||   |||   |||          |||      |||:
755 sProValCysProGluLysGlnArgSerArgAspLeuProSerLeuProA 772

600 CTGGACTCCGGGCCCCACCAAAGTCCTGCCAGCACAACGGGACCATGTAC 649
        ::              |||   ::::::|||       :::|||
772 sn......LeuGluProGlyGluGlyCysTyrPheAspGlyAspArgSer 786

650 CAACACGGAGAGATCTTCAGTGCCCATGAGCTGTTCCCCTCCCGCCTGCC 699
          ::::::                 ||.|   :::   |||
787 TrpArgAlaAlaGlyThrArgTrpHisProValValProProPheGlyLe 803

700 CAACCAGTGTGTCCTCTGCAGCTGC......ACAGAGGGCCAGATCTACT 743
      :::|||::::::|||:::|||      :::   |||:::::::::|
803 uIleLysCysAlaValCysThrCysLysGlyAlaThrGlyGluValHisC 820
```

FIG. 2C (CONT.[2])

```
744 GCGGCCTCACAACCTGCCCCGAACCAGGCTGCCCAGCACCCCTCCCGCTG 793
       ||       :::    ||||||:::    :::|||       |||:::
820 ysGluLysValGlnCysProArgLeuAlaCysAlaGlnProValArgAla 836

794 ...CCAGACTCCTGCTGCCAAGCCTGCAAAGATGAGGCAAGTGAGCAATC 840
                  |||    :::||||||:::    |||
837 AsnProThrAspCysCysLysGlnCys..................... 845

841 GGATGAAGAGGACAGTGTGCAGTCGCTCCATGGGGTGAGACATCCTCAGG 890

845 ..................................................... 845

891 ATCCATGTTCCAGTGATGCTGGGAGAAAGAGAGGCCCGGGCACCCCAGCC 940
                                    |||:::
846 .....................................ProVal 847

941 CCCACTGGCCTCAGCGCCCCTCTGAGCTTCATCCCTCGCCACTTCAGACC 990
      :::|||   :::|||   |||:::         :::          ||
848 GlySerGlyThrAsnAlaLysLeuGlyAsp.ProMetGlnAlaAspGlyPr 864

991 CAAGGGAGCAGGCAGCACAACTGTCAAGATCGTCCTGAAGGAGAAACATA 1040
                              |:::|||
864 oArgGly.................................... 866
```

FIG. 2C (CONT.[3])

```
1041 NGAAAGCCTGTGTGCATGGCGGGAAGACGTACTCCCACGGGGAGGTGTGG 1090
         |||      :::|||:::   :::   :::::::::   |||
 867 ........CysArgPheAlaGlyGlnTrpPheProGluAsnGlnSerTrp 880

1091 CACCCGGCCTTCCGTGCCTTCGGCCCTTGCCCATGCATCCTATGCACCTG 1140
     ||||||:::          ||||||          ||||||   |||   ||
 881 HisProSerValProProPheGlyGluMetSerCysIleThrCysArgCy 897

1141 TGAGGATGGCCGCCAGGACTGCCAGCGTGTGACCTGTCCCACGAAGTACC 1190
           |     |||          |||:::|||      |||
 897 sGlyAlaGlyValProHisCysGluArgAspAspCysSerProProLeuS 914

1191 CCTGCCGTCACCCCGAGAAAGTGGCTGGGAAGTGCTGCAAGATTTGC... 1237
       |||         ::::::      ::::::||||||::::   |||
 914 erCysGlySerGlyLysGlu......SerArgCysCysSerHisCysThr 928

1238 ..............................CCAGAGGACAAAGCAGACCC 1257
                      ||||||   :::   :::
 929 AlaGlnArgSerSerGluThrArgThrLeuProGluLeuGluLysGluAl 945

1258 TGGCCACAGT 1267
       ||||||
 945 aGluHisSer 948
```

FIG. 2C (CONT.[4])

```
656 TCGCTCCATGGGGTGAGACATCCTCAGGATCCATGTTCCAGTGATGCTGG 705
    |||   |||      ::::::          |||      :::   :::||
  3 SerTyrHisArgSerHisTyrAspProProProSerArgGlnAlaGlyGl 19

706 GAGAAAGAGAGGCCCGGGCACCCCAGCCCCCACTGGCCTCAGCGCCCCTC 755
          |   :::|||   ||||||:::   :::      |||
 19 yLeuSerArgPheProGlyAlaArgSerHisArgGlyAlaLeu..... 33

756 TGAGCTTCATCCCTCGCCACTTCAGACCCAAGGGAGCAGGCAGCACAACT 805
              :::      :::   |||:::::::|||
 34 .....MetAspSerGlnGlnAlaSerGlyThrIle 43

806 GTCAAGATCGTCCTG......AAGGAGAAACATANGAAAGCCTGTGTGCA 849
    |||:::|||||||:::      |||:::||||||   ::::::||||||
 44 ValGlnIleValIleAsnAsnLysHisLysHisGlyGlnValCysValSe 60

850 TGGCGGGAAGACGTACTCCCACGGGGAGGTGTGGCACCCGGCCTTCCGTG 899
    :::|||||||||||||||||||||||||  ||||||||   :::||||
 60 rAsnGlyLysThrTyrSerHisGlyGluSerTrpHisProAsnLeuArgA 77
```

FIG. 3A

```
 900 CCTTCGGCCCTTGCCCATGCATCCTATGCACCTGTGAGGATGGCCGCCAG 949
     ||||||||         |||:::|||||||||||||:::      :::|||
  77 laPheGlyIleValGluCysValLeuCysThrCysAsnValThrLysGln 93

 950 GACTGCCAGCGTGTGACCTGTCCCACGAAGTACCCCTGCCGTCACCCCGA 999
     :::|||:::::::::   ||||||::::::||||||||||::::::|||::
  94 GluCysLysLysIleHisCysProAsnArgTyrProCysLysTyrProGl 110

1000 GAAAGTGGCTGGGAAGTGCTGCAAGATTTGCCCAGAGGACAAAGCAGAC. 1048
       :|||:::   ||||||||||||||::||||||      ||||||
 110 nLysIleAspGlyLysCysCysLysValCysProGlyLysLysAlaLysG 127

1049 ........CCTGGCCACAGT...GAGATCAGTTCTACCAGGTGTCCCAAG 1087
         ||||||:::|||   :::   ::::::      |||   :::
 127 luGluLeuProGlyGlnSerPheAspAsnLysGlyTyrPheCysGlyGlu 143

1088 GCACCGGGCCGGGTCCTCGTCCACACATCGGTA...TCCCCAAGCCCAGA 1134
           :::   |||:::   ||||||          :::   ::
 144 Glu......ThrMetProValTyrGluSerValPheMetGluAspGlyGl 158
```

FIG. 3A (CONT.[1])

```
1135 CAACCTGCGTCGCTTTGCCCTGGAACACGAGGCCTCGGACTTGGTGGAGA 1184
     ::::   |||:::::::|||||||||   |||             ||||||:
 158 uThrThrArgLysIleAlaLeuGluThrGluArgProProGlnValGluV 175
1185 TCTACCTCTGGAAGCTGGTAAANNNNNNNNNNNNNNNNNNNNNNNNNNN 1234
                    :::::::|||   :::
 175 alHisValTrpThrIle..................................  180

1235 NNNNNNNNNNNNNNNNNNNNNNNNNNNCAGAATTTCCACTTGACTCAGAT 1284
                  |||:::||||||         ||
 181 ............ArgLysGlyIleLeuGlnHisPheHis.........Il 190

1285 CAAGAAAGTCAGGAAGCAAGACTTCCAGAAAGAGGCACAGCACTTCCGAC 1334
     |:::|||:::   |||:::   |||:::   |||       ||||||:::|
 190 eGluLysIleSerLysArgMetPheGlu...GluLeuProHisPheLysL 206

1335 TGCTCGCTGGCCCCCACGAAGGTCACTGGAACGTCTTCCTAGCCCAGACC 1384
           ||::::::              ::::::|||:::::::|||
 206 euValThrArgThrThrLeuSerGlnTrpLysIlePheThrGluGlyGlu 222

1385 CTGGAGCTGAAGGTCACGGCCAGTCCAGACAAAGTGACCAAGACATAACA 1434
       :::::::::      :::|||                      +++
 223 AlaGlnIleSerGlnMetCysSer.........................  230
```

FIG. 3A (CONT.[2])

```
1435 AAGACCTAACAGTTGCAGATATGAGCTGTATAATTGTTGTTATTATATAT 1484
         :::   ||||||   ||||||   :::              :::
 231 ....SerArgValCysArgThrGluLeuGluAspLeuValLysValLeuT 246

1485 TAATAAATAAGAAGTTGCATAACCAT 1510
     ::::::::::::      :::|||
 246 yrLeuGluArgSerGluLysGlyHis 254
```

FIG. 3A (CONT.[3])

```
368 CCCACTGTGGAACCTCACACTCCCTCTGGACTCCGGGCCCCA......CC 411
    |||::::::   |||      |||:::|||          |||      ||
532 ProAlaLeuAlaProAlaLysProGlyGlyProGlyArgProArgAspPr 548

412 AAAGTCCTGCCAGCACAACGGGACCATGTACCAACACGGAGAGATCTTCA 461
    |:::::::|||      :::|||   :::      ||||||      ::::
548 oAsnThrCysPhePheGluGlyGlnGlnArgProHisGlyAlaArgTrpA 565

462 GTGCCCATGAGCTGTTCCCCTCCCGCCTGCCCAACCAG.........TGT 502
    ::                              ||||||          |||
565 la.............................ProAsnTyrAspProLeuCys 572

503 GTCCTCTGCAGCTGCACAGAGGGCCAGATCTACTGCGGCCTCACAACCTG 552
     ||||||:::|||   :::      :::   |||      ::::::||
573 SerLeuCysThrCysGlnArgArgThrValIleCysAspProValValCy 589

553 CCCCGAACCAGGCTGCCCAGCACCCCTCCCGCTGCCAGACTCCTGCTGCC 602
    ||||   |||:::||||||   |||:::      ||||||:::|||||||
589 sProProProSerCysProHisProValGlnAlaProAspGlnCysCysP 606

603 AAGCCTGCAAAGATGAGGCAAGTGAGCAATCGGATGAAGAGGACAGTGTG 652
     :::|||                              |||::::::::::|||
606 roValCys.........................ProGluLysGlnAspVal 614
```

FIG. 3B

```
653 CAGTCGCTCCATGGGGTGAGACATCCTCAGGATCCA.........TGTTC 693
     :::::::|||   |||:::   :::   :::||||||          |||
615 ArgAspLeuProGlyLeuProArgSerArgAspProGlyGluGlyCysTy 631

694 CAGTGATGCTGGGAGAAAG...AGAGGCCCGGGCACC..............  727
          |||:::   |||:::   |||:::   ||||||
631 rPheAspGlyAspArgSerTrpArgAlaAlaGlyThrArgTrpHisProV 648

728 ..CCAGCCCCCACTGGCCTC.............................  745
                      |||   ||||||
648 alValProProPheGlyLeuIleLysCysAlaValCysThrCysLysGly 664

745 .............................................  745

665 GlyThrGlyGluValHisCysGluLysValGlnCysProArgLeuAlaCy 681

746 .AGCGCCCCTCTGAGCTTCATCCCTCGCCACTTC.........AGACCCA 785
     :::   |||:::         |||                    |||
681 sAlaGlnProValArgValAsnProThrAspCysCysLysGlnCysProV 698
```

FIG. 3B (CONT.[1])

```
786 AGGGAGCAGGCAGC...ACAACTGTCAAGATCGTCCTGAAGGAGAAACAT 832
        |||:::|||:::         :::          ::::::
698 alGlySerGlyAlaHisProGlnLeuGlyAspProMetGlnAlaAspGly 714

833 ANGAAAGCCTGTGTGCATGGCGGGAAGACGTACTCCCACGGGGAGGTGTG 882
     ::::::|||      :::|||:::   :::   :::::::::   ||
715 ProArgGlyCysArgPheAlaGlyGlnTrpPheProGluSerGlnSerTr 731

883 GCACCCGGCCTTCCGTGCCTTCGGCCCTTGCCCATGCATCCTATGCACCT 932
    |||||||:::         ||||||          ||||||   |||   |
731 pHisProSerValProProPheGlyGluMetSerCysIleThrCysArgC 748

933 GTGAGGATGGCCGCCAGGACTGCCAGCGTGTGACCTGTCCCACGAAGTAC 982
         ||     |||          |||:::|||      |||
748 ysGlyAlaGlyValProHisCysGluArgAspAspCysSerLeuProLeu 764

983 CCCTGCCGTCACCCCGAGAAAGTGGCTGGGAAGTGCTGCAAGATTTGC.. 1030
      |||         ::::::      ::::::|||||||:::   |||
765 SerCysGlySerGlyLysGlu......SerArgCysCysSerArgCysTh 779

1031 .....................CCAGAGGACAAAGCAGACCCT 1051
                  ||||||   ::::::||||||
 779 rAlaHisArgArgProAlaProGluThrArgThrAspPro 792
```

FIG. 3B (CONT.[2])

```
1031 TCGCTCCATGGGGTGAGACATCCTCAGGATCCATGTTCCAGTGATGCTGG 1080
     |||   |||      ::::::          |||      :::   :::||
   3 SerTyrHisArgSerHisTyrAspProProProSerArgGlnAlaGlyGl 19

1081 GAGAAAGAGAGGCCCGGGCACCCCAGCCCCCACTGGCCTCAGCGCCCCTC 1130
           |   :::|||   ||||||:::   :::      |||
  19 yLeuSerArgPheProGlyAlaArgSerHisArgGlyAlaLeu..... 33

1131 TGAGCTTCATCCCTCGCCACTTCAGACCCAAGGGAGCAGGCAGCACAACT 1180
               :::      :::   |||:::::::|||
  34 .....MetAspSerGlnGlnAlaSerGlyThrIle 43

1181 GTCAAGATCGTCCTG......AAGGAGAAACATANGAAAGCCTGTGTGCA 1224
      |||:::|||||||:::      |||:::||||||   ::::::||||||
  44 ValGlnIleValIleAsnAsnLysHisLysHisGlyGlnValCysValSe 60

1225 TGGCGGGAAGACGTACTCCCACGGGGAGGTGTGGCACCCGGCCTTCCGTG 1274
     :::|||||||||||||||||||||||||   |||||||||   :::||||
  60 rAsnGlyLysThrTyrSerHisGlyGluSerTrpHisProAsnLeuArgA 77
```

FIG. 4A

```
1275 CCTTCGGCCCTTGCCCATGCATCCTATGCACCTGTGAGGATGGCCGCCAG 1324
     ||||||||         |||:::|||||||||||||||:::      :::|||
  77 laPheGlyIleValGluCysValLeuCysThrCysAsnValThrLysGln 93

1325 GACTGCCAGCGTGTGACCTGTCCCACGAAGTACCCCTGCCGTCACCCCGA 1374
     :::|||:::::::::   ||||||:::::||||||||||:::::|||::
  94 GluCysLysLysIleHisCysProAsnArgTyrProCysLysTyrProGl 110

1375 GAAAGTGGCTGGGAAGTGCTGCAAGATTTGCCCAGAGGACAAAGCAGAC. 1423
       :|||:::   ||||||||||||||::||||||      ||||||
 110 nLysIleAspGlyLysCysCysLysValCysProGlyLysLysAlaLysG 127

1424 .........CCTGGCCACAGT...GAGATCAGTTCTACCAGGTGTCCCAAG 1462
         ||||||:::|||   :::   ::::::      |||   :::
 127 luGluLeuProGlyGlnSerPheAspAsnLysGlyTyrPheCysGlyGlu 143

1463 GCACCGGGCCGGGTCCTCGTCCACACATCGGTA...TCCCCAAGCCCAGA 1509
           :::   |||:::   ||||||          :::   ::
 144 Glu......ThrMetProValTyrGluSerValPheMetGluAspGlyGl 158
```

FIG. 4A (CONT.[1])

```
1510 CAACCTGCGTCGCTTTGCCCTGGAACACGAGGCCTCGGACTTGGTGGAGA 1559
     ::::   |||::::::|||||||||   |||              ||||||:
 158 uThrThrArgLysIleAlaLeuGluThrGluArgProProGlnValGluV 175

1560 TCTACCTCTGGAAGCTGGTAAANNNNNNNNNNNNNNNNNNNNNNNNNNN 1609
                     :::::::::|||   :::
 175 alHisValTrpThrIle..............................  180

1610 NNNNNNNNNNNNNNNNNNNNNNNNNNNCAGAATTTCCACTTGACTCAGAT 1659
                  |||:::||||||         ||
 181 ............ArgLysGlyIleLeuGlnHisPheHis.........Il 190

1660 CAAGAAAGTCAGGAAGCAAGACTTCCAGAAAGAGGCACAGCACTTCCGAC 1709
     |:::|||:::   |||:::   |||:::   |||      ||||||:::|
 190 eGluLysIleSerLysArgMetPheGlu...GluLeuProHisPheLysL 206

1710 TGCTCGCTGGCCCCCACGAAGGTCACTGGAACGTCTTCCTAGCCCAGACC 1759
           ||::::::              ::::::|||::::::|||
 206 euValThrArgThrThrLeuSerGlnTrpLysIlePheThrGluGlyGlu 222
```

FIG. 4A (CONT.[2])

```
1760 CTGGAGCTGAAGGTCACGGCCAGTCCAGACAAAGTGACCAAGACATAACA 1809
       :::::::::      :::|||                        +++
 223 AlaGlnIleSerGlnMetCysSer..........................  230

1810 AAGACCTAACAGTTGCAGATATGAGCTGTATAATTGTTGTTATTATATAT 1859
        :::   ||||||   ||||||   :::              :::
 231 ....SerArgValCysArgThrGluLeuGluAspLeuValLysValLeuT 246

          1860 TAATAAATAAGAAGTTGCATAACCAT 1885
               :::::::::::      :::|||
           246 yrLeuGluArgSerGluLysGlyHis 254
```

FIG. 4A (CONT.[3])

```
560 TGCCTTTTCCATGGGAAGAGATACTCCCCCGGCGAGAGCTGGCACCCCTA 609
    |||:::|||:::||:|:::::::::::::   |||::::::|||   |||
691 CysPhePheGluGlyGluGlnHisThrHisGlySerGlnTrpThrProGl 707

610 CTTGGAGCCACAAGGCCTGATGTACTGCCTGCGCTGTACCTGCTCAGAGG 659
      :::                      |||:::   |||||||||::::::
707 nTyrAsnThr.................CysPheThrCysThrCysGlnLysL 719

660 GCGCCCATGTGAGTTGTTACCGCCTCCACTGTCCGCCTGTCCACTGCCCC 709
      :::   |||   |||      :::   ||||||   :::   |||
719 ysThr...ValIleCysAspProValMetCysProThrLeuSerCysThr 734

710 CAGCCTGTGACGGAGCCACAGCAATGCTGTCCCAAGTGTGTGGAACCTCA 759
      :::   |||         :::||||||||||||   |||   |||
735 HisThrValGlnProGluAspGlnCysCysProIleCysGluGluLysLy 751

760 CACTCCCTCTGGACTCCGGGCC.....................CCACCAAAGT 791
      :::         |||              .            |||::::
751 sGluSerLysGluThrAlaAlaValGluLysValGluGluAsnProGluG 768
```

FIG. 4B

```
792 CCTGCCAGCACAACGGGACCATGTACCAACACGGAGAGATCTTCAGTGCC 841
      ::|||       :::|||   :::   :::|||:::           :::
768 lyCysTyrPheGluGlyAspGlnLysMetHisAlaProGlyThrThrTrp 784

842 CATGAGCTGTTCCCCTCCCGCCTGCCCAACCAGTGTGTCCTCTGCAGCTG 891
    |||   :::   |||                 :::|||::::::|||:::||
785 HisProPheValProProPheGlyTyrIleLysCysAlaValCysThrCy 801

892 C......ACAGAGGGCCAGATCTACTGCGGCCTCACAACCTGCCCCGAAC 935
      |      :::   |||::::::::::|||      :::||||||||||
801 sLysGlySerThrGlyGluValHisCysGluLysValThrCysProProL 818

936 CAGGCTGCCCAGCACCCCTCCCGCTG...CCAGACTCCTGCTGCCAAGCC 982
        |||      |||:::          |||::::::|||||||:::
818 euThrCysSerArgProIleArgArgAsnProSerAspCysCysLysGlu 834

983 TGCAAAGATGAGGCAAGTGAGCAATCGGATGAAGAGGACAGTGTGCAGTC 1032
    |||      |||   :::          ::::::|||:::   :::|||::
835 CysProProGluGluThrProProLeuGluAspGluGluMetMetGlnAl 851
```

FIG. 4B (CONT.[1])

```
1033 GCTCCATGGGGTGAGACATCCTCAGGATCCATGTTCCAGTGATGCTGGGA 1082
                                  :
 851 a..................................................  851

1083 GAAAGAGAGGCCCGGGCACCCCAGCCCCCACTGGCCTCAGCGCCCCTCTG 1132
                                |||||||
 852 ..........AspGlyThr.....................................  854

1133 AGCTTCATCCCTCGCCACTTCAGACCCAAGGGAGCAGGCAGCACAACTGT 1182

 854 ..................................................  854

1183 CAAGATCGTCCTGAAGGAGAAACATANGAAAGCCTGTGTGCATGGCGGGA 1232
                    :::   |||      |||   :
 855 ............................ArgLeuCysLysPheGlyLysA 862

1233 AGACGTACTCCCACGGGGAGGTGTGGCACCCGGCCTTCCGTGCCTTCGGC 1282
     ::   |||::::::::::|||   |||||||||::::            |||
 862 snTyrTyrGlnAsnSerGluHisTrpHisProSerValProLeuValGly 878
```

FIG. 4B (CONT.[2])

```
1283 CCTTGCCCATGCATCCTATGCACCTGTGAGGATGGCCGCCAGGACTGCCA 1332
          ||||||   |||   |||:::   |||         |||||
 879 GluMetLysCysIleThrCysTrpCysAspHisGlyValThrLysCysGl 895

1333 GCGTGTGACCTGTCCCACGAAGTACCCCTGCCGTCACCCCGAGAAAGTGG 1382
      ||||      ||||||            ||||||:::|||   ::::::
 895 nArgLysGlnCysProLeu...LeuSerCysArgAsnProIleArgThrG 911

         1383 CTGGGAAGTGCTGCAAGATTTGCCCAGAGGAC 1414
               ||||||||||||      |||   ||||||
       911 luGlyLysCysCysProGluCysIleGluAsp
```

FIG. 4B (CONT.[3])

```
560 TGCCTTTTCCATGGGAAGAGATACTCCCCCGGCGAGAGCTGGCACCCCTA 609
     |||   |||:::|||:::      :::   |||:::   |||   |||
693 CysSerPheGluGlyGlnLeuArgAlaHisGlySerArgTrpAlaProAs 709

610 CTTGGAGCCACAAGGCCTGATGTACTGCCTGCGCTGTACCTGCTCAGAGG 659
      :::   :::             |||      |||:::|||::::::
709 pTyrAspArgLys..........CysSerValCysSerCysGlnLysA 722

660 GCGCCCATGTGAGTTGTTACCGCCTCCACTGTCCGCCTGTCCACTGCCCC 709
      :::   |||   |||      :::   |||||||||::::::|||
722 rgThr...ValIleCysAspProIleValCysProProLeuAsnCysSer 737

710 CAGCCTGTGACGGAGCCACAGCAATGCTGTCCCAAGTGTGTGGAACCTCA 759
       |||||||||      |||:::||||||||||||   |||   |||
738 GlnProValHisLeuProAspGlnCysCysProValCysGluGluLysLy 754

760 CACTCCCTCTGGACTCCGGGCCCCACCAAAGTCCTGCCAGCACAACGGGA 809
          ::::::   |||   ::::::          :::|||
754 sGluMetArgGluValLysLysProGluArgAlaArgThrSerGluGlyC 771
```

FIG. 4C

```
810 CCATGTACCAACACGGAGAGATCTTCAGTGCC.............CAT 844
     :::::::::      :::   :::::::|||                |||
771 ysPhePheAspGlyAspArgSerTrpLysAlaAlaGlyThrArgTrpHis 787

845 GAGCTGTTCCCCTCCCGCCTGCCCAACCAGTGTGTCCTCTGCAGCTGC.. 892
      :::   |||                  :::|||::::::|||:::|||
788 ProPheValProProPheGlyLeuIleLysCysAlaIleCysThrCysLy 804

893 ....ACAGAGGGCCAGATCTACTGCGGCCTCACAACCTGCCCCGAACCAG 938
      :::   |||::::::::::|||      :::|||||||||:::   :
804 sGlySerThrGlyGluValHisCysGluLysValThrCysProLysLeuS 821

939 GCTGCCCAGCACCCCTCCCGCTG...CCAGACTCCTGCTGCCAAGCCTGC 985
    ::|||      |||:::          |||::::::||||||:::   |||
821 erCysThrAsnProIleArgAlaAsnProSerAspCysCysLysGlnCys 837

986 AAAGATGAGGCAAGTGAGCAATCGGATGAAGAGGACAGTGTGCAGTCGCT 1035
        |||      :::      :::    . ||||||:::||||||
838 ProValGluGluArgSerProMetGluLeuAlaAspSerMetGlnSer.. 853
```

FIG. 4C (CONT.[1])

```
1036 CCATGGGGTGAGACATCCTCAGGATCCATGTTCCAGTGATGCTGGGAGAA 1085

 853 ..................................................  853

1086 AGAGAGGCCCGGGCACCCCAGCCCCCACTGGCCTCAGCGCCCCTCTGAGC 1135

 853 ..................................................  853

1136 TTCATCCCTCGCCACTTCAGACCCAAGGGAGCAGGCAGCACAACTGTCAA 1185
                          ||||||||||||
 854 ........................AspGlyAlaGlySer..........  858

1186 GATCGTCCTGAAGGAGAAACATANGAAAGCCTGTGTGCATGGCGGGAAGA 1235
                            |||      ||||
 859 ..............................CysArgPheGlyArgHisT  865

1236 CGTACTCCCACGGGGAGGTGTGGCACCCGGCCTTCCGTGCCTTCGGCCCT 1285
       |||   :::   |||   |||||||||:::          ||||||
 865 rpTyrProAsnHisGluArgTrpHisProThrValProProPheGlyGlu  881

1286 TGCCCATGCATCCTATGCACCTGTGAGGATGGC.................CG 1320
       |||:::   |||||||||   :::|||                    ||
 882 MetLysCysValThrCysThrCysAlaGluGlyIleThrGlnCysArgAr  898
```

FIG. 4C (CONT.[2])

```
1321 CCAGGACTGCCAGCGTGTGACCTGTCCCACGAAGTACCCCTGCCGTCACC 1370
              ||||:::|||      :::||||||   |||
 898 gGlnGluCysThrGlyThrThrCysGlyThr..................... 908

1371 CCGAGAAAGTGGCTGGGAAGTGCTGCAAGATTTGCCCAGAGGACAAAGCA 1420
        :::          :::||||||      |||   :::   :::
 909 ..GlySerLysArgAspArgCysCysThrLysCysLysAspAlaAsnGln 924

   1421 GACCCTGGCCACAGT...GAGATCAGTTCTACCAGGTGTCCC 1459
        |||      ::::::   :::   ::::::||||||   |||
    925 AspGluAspGluLysValLysSerAspGluThrArgThrPro 938
```

FIG. 4C (CONT.[3])

```
536 GGTTTGCCAGGCCCAGACATGTTCTGCCTTTTCCATGGGAAGAGATACTC 585
      |||   |||                   |||:::|||:::|||:::::: 
543 GlyArgProArgAspProAsnThrCysPhePheGluGlyGlnGlnArgPr 559

586 CCCCGGCGAGAGCTGGCACCCCTACTTGGAGCCACAAGGCCTGATGTACT 635
      |||       |||   |||       :::|||                 |
559 oHisGlyAlaArgTrpAlaProAsnTyrAspPro.............LeuC 572

636 GCCTGCGCTGTACCTGCTCAGAGGGCGCCCATGTGAGTTGTTACCGCCTC 685
    ||      |||||||||:::::   :::   |||   |||      ::: 
572 ysSerLeuCysThrCysGlnArgArgThr...ValIleCysAspProVal 587

686 CACTGTCCGCCTGTCCACTGCCCCCAGCCTGTGACGGAGCCACAGCAATG 735
      |||||||||      ||||||:::||||||      |||:::||||| 
588 ValCysProProProSerCysProHisProValGlnAlaProAspGlnCy 604

736 CTGTCCCAAGTGTGTGGAA.....................CCTCACACTC 764
    |||||||   |||   |||                     |||      | 
604 sCysProValCysProGluLysGlnAspValArgAspLeuProGlyLeuP 621
```

FIG. 4D

```
765 CCTCTGGACTCCGGGCCCCACCAAAGTCCTGCCAGCACAACGGGACCATG 814
      ||   :::   |||   |||   :::::::|||       :::|||
621 roArgSer...ArgAspProGlyGluGlyCysTyrPheAspGlyAspArg 636

815 TACCAACACGGAGAGATCTTCAGTGCCCATGAGCTGTTCCCCTCCCGCCT 864
        ::::::                  |||   :::   |||
637 SerTrpArgAlaAlaGlyThrArgTrpHisProValValProProPheGl 653

865 GCCCAACCAGTGTGTCCTCTGCAGCTGCACA......GAGGGCCAGATCT 908
      :::|||::::::|||:::|||              |||:::::::
653 yLeuIleLysCysAlaValCysThrCysLysGlyGlyThrGlyGluValH 670

909 ACTGCGGCCTCACAACCTGCCCCGAACCAGGCTGCCCAGCACCCCTCCCG 958
     ::|||     :::   ||||||:::   :::|||       |||:::
670 isCysGluLysValGlnCysProArgLeuAlaCysAlaGlnProValArg 686

959 CTG...CCAGACTCCTGCTGCCAAGCCTGCAAAGATGAGGCAAGTGAGCA 1005
             :::   |||   :::|||||||:::   |||
687 ValAsnProThrAspCysCysLysGlnCys..................... 696
```

FIG. 4D (CONT.[1])

```
1006 ATCGGATGAAGAGGACAGTGTGCAGTCGCTCCATGGGGTGAGACATCCTC 1055
                  :::   :::      |||||||
 697 ........................ProValGlySerGlyAlaHisProG 705

1056 AG......GATCCATGTTCCAGTGATGCTGGGAGAAAGAGAGGCCCGGGC 1099
       ||      ||||||   :::::::|||                ||||||
 705 lnLeuGlyAspProMetGlnAlaAsp..............GlyPro... 715

1100 ACCCCAGCCCCCACTGGCCTCAGCGCCCCTCTGAGCTTCATCCCTCGCCA 1149

 715 ..................................................  715

1150 CTTCAGACCCAAGGGAGCAGGCAGCACAACTGTCAAGATCGTCCTGAAGG 1199

 715 ..................................................  715

1200 AGAAACATANGAAAGCCTGTGTGCATGGCGGGAAGACGTACTCCCACGGG 1249
          :::::::|||      :::|||:::   :::   ::::::
 716 ...........ArgGlyCysArgPheAlaGlyGlnTrpPheProGluSer 728

1250 GAGGTGTGGCACCCGGCCTTCCGTGCCTTCGGCCCTTGCCCATGCATCCT 1299
       :::   |||||||||:::          ||||||          ||||||
 729 GlnSerTrpHisProSerValProProPheGlyGluMetSerCysIleTh 745
```

FIG. 4D (CONT.[2])

```
1300 ATGCACCTGTGAGGATGGCCGCCAGGACTGCCAGCGTGTGACCTGTCCCA 1349
       |||   |||       |||          |||:::|||       |||
 745 rCysArgCysGlyAlaGlyValProHisCysGluArgAspAspCysSerL 762

1350 CGAAGTACCCCTGCCGTCACCCCGAGAAAGTGGCTGGGAAGTGCTGCAAG 1399
         |||          ::::::      :::::::::::::::
 762 euProLeuSerCysGlySerGlyLysGlu......SerArgCysCysSer 776

 1400 ATTTGC.....................CCAGAGGACAAAGCAGACCCT 1426
       |||                      ||||||   ::::::||||||
  777 ArgCysThrAlaHisArgArgProAlaProGluThrArgThrAspPro 792
```

FIG. 4D (CONT.[3])

```
chordin_ed7       SPLPSAGPSF VSPSLPPFPA FSFHLSLLPT LDLPSCPPFL PTAASWPFSD
chordin_ed6TR_2   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
chordin_ed6TR_1   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

                  51                                                  100
    chordin_ed7   PALAADLLGS CGLICGPCXS VSFSSPVLPT PLPDQRPDPG ERMVPEVRVL
chordin_ed6TR_2   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
chordin_ed6TR_1   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

                  101                                                 150
    chordin_ed7   SSLLGLALLW FPLDSHARAR PDMFCLFHGK RYSPGESWHP YLEPQGLMYC
chordin_ed6TR_2   DRVFGLEPPG TNMALVGLPG PDMFCLFHGK RYSPGESWHP YLEPQGLMYC
chordin_ed6TR_1   ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~

                  151                                                 200
    chordin_ed7   LRCTCSEGAH VSCYRLHCPP VHCPQPVTE. PQQCCPK.CV EPHTPSGLRA
chordin_ed6TR_2   LRCTCSEGAH VSCYRLHCPP VHCPQPVTE. PQQCCPK.CV EPHTPSGLRA
chordin_ed6TR_1   ~~~ISSWGQM QNHQKSGLVN FSKDSHETSF SSSSCPSPTV EPHTPSGLRA
```

FIG. 5

```
                201                                                 250
    chordin_ed7 PPKSCQHNGT MYQHGEIFSA HELFPSRLPN QCVLCSCTEG QIYCGLTTCP
chordin_ed6TR_2 PPKSCQHNGT MYQHGEIFSA HELFPSRLPN QCVLCSCTEG QIYCGLTTCP
chordin_ed6TR_1 PPKSCQHNGT MYQHGEIFSA HELFPSRLPN QCVLCSCTEG QIYCGLTTCP

                251                                                 300
    chordin_ed7 EPGCPAPLPL PDSCCQACKD EASEQSDEED SVQSLHGVRH PQDPCSSDAG
chordin_ed6TR_2 EPGCPAPLPL PDSCCQACKD EASEQSDEED SVQSLHGVRH PQDPCSSDAG
chordin_ed6TR_1 EPGCPAPLPL PDSCCQACKD EASEQSDEED SVQSLHGVRH PQDPCSSDAG

                301                                                 350
    chordin_ed7 RKRGPGTPAP TGLSAPLSFI PRHFRPKGAG STTVKIVLKE KHXKACVHGG
chordin_ed6TR_2 RKRGPGTPAP TGLSAPLSFI PRHFRPKGAG STTVKIVLKE KHXKACVHGG
chordin_ed6TR_1 RKRGPGTPAP TGLSAPLSFI PRHFRPKGAG STTVKIVLKE KHXKACVHGG

                351                                                 400
    chordin_ed7 KTYSHGEVWH PAFRAFGPCP CILCTCEDGR QDCQRVTCPT KYPCRHPEKV
chordin_ed6TR_2 KTYSHGEVWH PAFRAFGPCP CILCTCEDGR QDCQRVTCPT KYPCRHPEKV
chordin_ed6TR_1 KTYSHGEVWH PAFRAFGPCP CILCTCEDGR QDCQRVTCPT KYPCRHPEKV
```

FIG. 5 (CONT.[1])

```
                401                                                 450
    chordin_ed7 AGKCCKICPE DKADPGHSEI SSTRCPKAPG RVLVHTSVSP SPDNLRRFAL
chordin_ed6TR_2 AGKCCKICPE DKADPGHSEI SSTRCPKAPG RVLVHTSVSP SPDNLRRFAL
chordin_ed6TR_1 AGKCCKICPE DKADPGHSEI SSTRCPKAPG RVLVHTSVSP SPDNLRRFAL

                451                                                 500
    chordin_ed7 EHEASDLVEI YLWKLVKDEE TEAQRGEVPG PRPHSQNFHL TQIKKVRKQD
chordin_ed6TR_2 EHEASDLVEI YLWKLVKDEE TEAQRGEVPG PRPHSQNFHL TQIKKVRKQD
chordin_ed6TR_1 EHEASDLVEI YLWKLVKDEE TEAQRGEVPG PRPHSQNFHL TQIKKVRKQD

                501                                             542
    chordin_ed7 FQKEAQHFRL LAGPHEGHWN VFLAQTLELK VTASPDKVTK T*
chordin_ed6TR_2 FQKEAQHFRL LAGPHEGHWN VFLAQTLELK VTASPDKVTK T*
chordin_ed6TR_1 FQKEAQHFRL LAGPHEGHWN VFLAQTLELK VTASPDKVTK T*
```

FIG. 5 (CONT.[2])

```
862 TCGCTCCATGGGGTGAGACATCCTCAGGATCCATGTTCCAGTGATGCTGG 911
    |||   |||      ::::::          |||      :::   :::||
  3 SerTyrHisArgSerHisTyrAspProProProSerArgGlnAlaGlyGl 19

912 GAGAAAGAGAGGCCCGGGCACCCCAGCCCCCACTGGCCTCAGCGCCCCTC 961
          |   :::|||   ||||||:::   :::      |||
 19 yLeuSerArgPheProGlyAlaArgSerHisArgGlyAlaLeuMetAsp. 35

962 TGAGCTTCATCCCTCGCCACTTCATACCCAAGGGAGCAGGCAGCACAACT 1011
                  :::   |||::::::|||
 36 ........................SerGlnGlnAlaSerGlyThrIle 43

1012 GTCAAGATCGTCCTGAAGGAGAAACATAAG......AAAGCCTGTGTGCA 1055
     |||:::||||||:::::::::|||||||      ::::::||||||
  44 ValGlnIleValIleAsnAsnLysHisLysHisGlyGlnValCysValSe 60

1056 TGGCGGGAAGACGTACTCCCACGGGGAGGTGTGGCACCCGGCCTTCCGTG 1105
     :::|||||||||||||||||||||||| ·|||||||||   :::||||
  60 rAsnGlyLysThrTyrSerHisGlyGluSerTrpHisProAsnLeuArgA 77
```

FIG. 6

```
1106 CCTTCGGCCCCTTGCCCTGCATCCTATGCACCTGTGAGGATGGCCGCCAG 1155
     ||||||||   :::   |||:::|||||||||||||:::      :::|||
  77 laPheGlyIleValGluCysValLeuCysThrCysAsnValThrLysGln 93

1156 GACTGCCAGCGTGTGACCTGTCCCACCGAGTACCCCTGCCGTCACCCCGA 1205
     :::|||:::::::::   |||||:::::|||||||||:::::|||::
  94 GluCysLysLysIleHisCysProAsnArgTyrProCysLysTyrProGl 110

1206 GAAAGTGGCTGGGAAGTGCTGCAAGATTTGCCCAGAGGACAAAGCAGAC. 1254
       :|||:::   |||||||||||||||::||||||      ||||||
 110 nLysIleAspGlyLysCysCysLysValCysProGlyLysLysAlaLysG 127

1255 ........CCTGGCCACAGT...GAGATCAGTTCTACCAGGTGTCCCAAG 1293
         ||||||:::|||   :::   ::::::      |||   :::
 127 luGluLeuProGlyGlnSerPheAspAsnLysGlyTyrPheCysGlyGlu 143

1294 GCACCGGGCCGGGTCCTCGTCCACACATCGGTA...TCCCCAAGCCCAGA 1340
           :::   |||:::   ||||||   .      :::   ::
 144 Glu......ThrMetProValTyrGluSerValPheMetGluAspGlyGl 158
```

FIG. 6 (CONT.[1])

```
1341 CAACCTGCGTCGCTTTGCCCTGGAACACGAGGCCTCGGACCTGGTGGAGA 1390
     ::::   |||::::::|||||||||   |||              ||||||:
 158 uThrThrArgLysIleAlaLeuGluThrGluArgProProGlnValGluV 175

1391 TCTACCTCTGGAAGCTGGTAAANNNNNNNNNNNNNNNNNNNNNNNNNNN 1440
                         :::::::::|||   :::
 175 alHisValTrpThrIle..................................  180

1441 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNCAGAATCTTCCACTTGACTCAGA 1490
                      |||::: ||||||          |
 181 ............ArgLysGlyIleLeuGlnHis.PheHis.........I 190

1491 TCAAGAAAGTCAGGAAGCAAGACTTCCAGAAAGAGGCACAGCACTTCCGA 1540
     ||:::|||:::   |||:::   |||:::   |||      ||||||:::
 190 leGluLysIleSerLysArgMetPheGlu...GluLeuProHisPheLys 205

1541 CTGCTCGCTGGCCCCCACGAAGGTCACTGGAACGTCTTCCTAGCCCAGAC 1590
           |||::::::                 :::::::|||:::::::|||
 206 LeuValThrArgThrThrLeuSerGlnTrpLysIlePheThrGluGlyGl 222
```

FIG. 6 (CONT.^{2})

```
1591 CCTGGAGCTGAAGGTCACGGCCAGTCCAGACAAAGTGACCAAGACATAAC 1640
                                                    +++
       :::::::::      :::|||                          .... 230
 222 uAlaGlnIleSerGlnMetCysSer.....

1641 AAAGACCTAACAGTTGCAGATATGAGCTGTATAATTGCTGTTATTATATA 1690
         :::   ||||||  ||||||   :::                 :::
 231 .....SerArgValCysArgThrGluLeuGluAspLeuValLysValLeu 245

1691 TTAATAAATAAGAAGTTGCATAACCAT 1717
           ::::::::::::      :::|||
 246 TyrLeuGluArgSerGluLysGlyHis 254
```

FIG. 6 (CONT.[3])

```
  3 CTTCCCCCTTTCTTTGATCGCCTCTCC.............CTTCTGCTGGA 40
    |||||||||      ::::::   :::              |||   |||||
390 LeuProProValLysSerGlnAlaAlaGlyHisAlaTrpLeuSerLeuAs 406

 41 CCTTCCTTCGTCTCTCCATCTCTCCCTCCTT..................... 71
               |              ||||||   :::::|||
406 pThrHisCysHisLeuHisTyrGluValLeuLeuAlaGlyLeuGlyGlyS 423

 72 ..TCCCCGCGTTCTCTTTCCACCTTTCTCTTCTTCCCACCTTAGACCTCC 119
      :::      ::::::::::::   |||:::   ||||||+++|||
423 erGluGlnGlyThrValThrAlaHisLeuLeuGlyProProGlyThr... 438

120 CTTCCTGCCCTCCTTTCCTGCCCACCGCTGCTTCCTGGCCCTTCTCCGAC 169
                          |||:::|||
439 ..................................ProGlyProAr 442

170 CCCGCTCTAGCAG.............CAGACCTCCTGGGGTCTGTGGGTTG 207
      |||+++:::                :::   :::   |||:::      |
442 gArgLeuLeuLysGlyPheTyrGlySerGluAlaGlnGlyValValLysA 459
```

FIG. 7

```
208 ATCTGTGGCCCCTGTGCCTCCGTGTCCTTTTCGTCTCCCTTCCTCCCGAC 257
              |||||   |||      ||||||   |||
459 spLeuGluProGluLeuLeuArgHisLeu..................... 468

258 TCCGCTCCCGGACCAGCGGCCTGACCCTGGGGAAAGGATGGTTCC...CG 304
     |||   |||   |||:::+++                    ||||||   ||
469 ...AlaLysGlyMetAlaSerLeuLeuIleThrThrLysGlySerProAr 484

305 AGGTGAG.....................................GGTC 315
     |||||||                                     |||
484 gGlyGluLeuArgGlyGlnValHisIleAlaAsnGlnCysGluValGlyG 501

316 CTCTCCTCCTT..............GCTGGGACTCGCGCT..... 341
          |||   |||                  |||:::||||||
501 lyLeuArgLeuGluAlaAlaGlyAlaGluGlyValArgAlaLeuGlyAla 517

342 ..........................................GCTCT 346
                              |||||
518 ProAspProAlaSerAlaAlaProProValValProGlyLeuProAlaLe 534
```

FIG. 7 (CONT.[1])

```
347 GGTTCCC......CCTGGACTCCCACGCTCGAGCCCGCCCAGACATGTTC 390
          |:::|||      ||||||   ||| :::|||   |||
534 uAlaProAlaLysProGlyGlyPro.GlyArgProArgAspProAsnThr 550

391 TGCCTTTTCCATGGGAAGAGATACTCCCCCGGCGAGAGCTGGCACCCCTA 440
     |||:::|||:::|||:::::          |||      |||   |||
551 CysPhePheGluGlyGlnGlnArgProHisGlyAlaArgTrpAlaProAs 567

441 CTTGGAGCCACAAGGCCTGATGTACTGCCTGCGCTGTACCTGCTCAGAGG 490
      :::|||                  |||      |||||||||:::::
567 nTyrAspPro.............LeuCysSerLeuCysThrCysGlnArgA 580

491 GCGCCCATGTGAGTTGTTACCGCCTCCACTGTCCGCCTGTCCACTGCCCC 540
     :::   |||   |||      :::   |||||||||      ||||||
580 rgThr...ValIleCysAspProValValCysProProProSerCysPro 595

541 CAGCCTGTGACGGAGCCACAGCAATGCTGTCCCAAGTGTGTGGAA..... 585
      :::||||||      |||:::|||||||||||||   |||   |||
596 HisProValGlnAlaProAspGlnCysCysProValCysProGluLysGl 612
```

FIG. 7 (CONT.^{2})

```
586 ..................CCTCACACTCCCTCTGGACTCCGGGCCCCACCAA 619
                 |||      |||   :::   |||   |||   :
612 nAspValArgAspLeuProGlyLeuProArgSer...ArgAspProGlyG 628

620 AGTCCTGCCAGCACAACGGGACCATGTACCAACACGGAGAGATCTTCAGT 669
          :::::|||      :::|||                :::::
628 luGlyCysTyrPheAspGlyAspArgSerTrpArgAlaAlaGlyThrArg 644

670 GCCCATGAGCTGTTCCCCTCCCGCCTGCCCAACCAGTGTGTCCTCTGCAG 719
     |||   :::   |||                    :::|||::::::|||::
645 TrpHisProValValProProPheGlyLeuIleLysCysAlaValCysTh 661

720 CTGCACA......GAGGGCCAGATCTACTGCGGGCTCACAACCTGCCCCG 763
    :|||           |||::::::::::|||    :::   ||||||:
661 rCysLysGlyGlyThrGlyGluValHisCysGluLysValGlnCysProA 678

764 AACCAGGCTGCCCAGCACCCCTCCCGCTG...CCAGACTCCTGCTGCCAG 810
    ::   :::|||     |||:::   ::: .  |||   :::||||||:::
678 rgLeuAlaCysAlaGlnProValArgValAsnProThrAspCysCysLys 694
```

FIG. 7 (CONT.[3])

```
811 GCCTGCAAAGATGAGGCAAGTGAGCAATCGGATGAAGAGGACCGTGTGCA 860
                              |||
695 GlnCys..................................... 696

861 GTCGCTCCATGGGGTGAGACATCCTCAG......GATCCATGTTCCAGTG 904
      :::   :::      |||||||||       ||||||   ::::::|
697 .ProValGlySerGlyAlaHisProGlnLeuGlyAspProMetGlnAlaA 713

905 ATGCTGGGAGAAAGAGAGGCCCGGGCACCCCAGCCCCCACTGGCCTCAGC 954
                    ||                  ||||||
713 sp.............GlyPro..................... 715

955 GCCCCTCTGAGCTTCATCCCTCGCCACTTCATACCCAAGGGAGCAGGCAG 1004

715 ..................................... 715

1005 CACAACTGTCAAGATCGTCCTGAAGGAGAAACATAAGAAAGCCTGTGTGC 1054
                              ::::::|||
716 ..............................ArgGlyCysArgP 720
```

FIG. 7 (CONT.[4])

```
1055 ATGGCGGGAAGACGTACTCCCACGGGGAGGTGTGGCACCCGGCCTTCCGT 1104
         :::|||:::   :::   :::::::::   |||||||||:::
 720 heAlaGlyGlnTrpPheProGluSerGlnSerTrpHisProSerValPro 736

1105 GCCTTCGGCCCCTTGCCCTGCATCCTATGCACCTGTGAGGATGGCCGCCA 1154
         ||||||   :::   ||||||   |||   |||       |||
 737 ProPheGlyGluMetSerCysIleThrCysArgCysGlyAlaGlyValPr 753

1155 GGACTGCCAGCGTGTGACCTGTCCCACCGAGTACCCCTGCCGTCACCCCG 1204
       |||:::|||      |||                |||         :
 753 oHisCysGluArgAspAspCysSerLeuProLeuSerCysGlySerGlyL 770

1205 AGAAAGTGGCTGGGAAGTGCTGCAAGATTTGC..................  1236
               :::::      :::::|||||||:::   |||
 770 ysGlu......SerArgCysCysSerArgCysThrAlaHisArgArgPro 784

1237 ...CCAGAGGACAAAGCAGACCCT 1257
        ||||||   ::::::|||||
 785 AlaProGluThrArgThrAspPro 792
```

FIG. 7 (CONT.[5])

```
  3 CTTCCCCCTTTCTTTGATCGCCTCTCC.............CTTCTGCTGGA 40
    |||||||||      ::::::   :::              |||   |||||
540 LeuProProValArgSerGlnAlaAlaGlyHisAlaTrpLeuSerLeuAs 556

 41 CCTTCCTTCGTCTCTCCATCTCTCCCTCCTT...................T 72
    |             ||||||   :::::::|||                   |
556 pThrHisCysHisLeuHisTyrGluValLeuLeuAlaGlyLeuGlyGlyS 573

 73 CCCCGCGTTCTCTTTCCACCTTTCTCTTCTTCCCACCTTAGACCTCCCTT 122
          ||   ::::::      |||                    +++
573 erGluGlnGlyThrValThr..............................  579

123 CCTGCCCTCCTTTCCTGCCCACCGCTGCTTCCTGGCCCTTCTCCGACCCC 172
           |||   |||      ||||||   :::||||||||||
580 ...AlaHisLeuLeuGlyProProGlyMetProGlyPro...........  591

173 GCTCTAGCAGCAG.........................ACCTCCTGGGGTC 198
      +++|||:::                          :::   |||:
592 .......GlnArgLeuLeuLysGlyPheTyrGlySerGluAlaGlnGlyV 606
```

FIG. 8

```
199 TGTGGGTTGATCTGTGGCCCCTGTGCCTCCGTGTCCTTTTCGTCTCCCTT 248
          ::      ||||||   |||:::   ||||||   |||
606 alValLysAspLeuGluProValLeuLeuArgHisLeu..... 618

249 CCTCCCGACTCCGCTCCCGGACCAGCGGCC..... 278
                    |||   |||   |||:::
619 .............AlaGlnGlyThrAlaSerLeuLeuIleThrThrLysSe 631

278 ..... 278

631 rSerProArgGlyGluLeuArgGlyGlnValHisIleAlaSerGlnCysG 648

279 .....TGACCCTGGGGA 290
                      +++|||   :::
648 luAlaGlyGlyLeuArgLeuAlaSerGluGlyValGlnMetProLeuAla 664

291 AAGGATGGTTCCCGAGGTGAGGGTCCTCTCCTCCTTGCTGGGACT...CG 337
          :::|||:::   :::   :::|||:::|||   ||||||
665 ProAsnGlyGluAlaAlaThrSerProMetLeuProAlaGlyProGlyPr 681
```

FIG. 8 (CONT.[1])

```
338 CGCTGCTCTGGTTCCCCCT......GGACTCCCACGCTCGAGCCCGCCCA 381
      |||   ||||||          |||   ||| :::|||   |||
681 oGluAlaProValProAlaLysHisGlySerPro.GlyArgProArgAsp 697

382 GACATGTTCTGCCTTTTCCATGGGAAGAGATACTCCCCCGGCGAGAGCTG 431
     |||:::|||:::|||:::::          |||      ||
698 ProAsnThrCysPhePheGluGlyGlnGlnArgProHisGlyAlaArgTr 714

432 GCACCCCTACTTGGAGCCACAAGGCCTGATGTACTGCCTGCGCTGTACCT 481
    |   |||      :::|||                |||      |||   |
714 pAlaProAsnTyrAspPro.............LeuCysSerLeuCysIleC 727

482 GCTCAGAGGGCGCCCATGTGAGTTGTTACCGCCTCCACTGTCCGCCTGTC 531
     ||::::::   :::   |||   |||      :::   |||||||||
727 ysGlnArgArgThr...ValIleCysAspProValValCysProProPro 742

532 CACTGCCCCCAGCCTGTGACGGAGCCACAGCAATGCTGTCCCAAGTGTGT 581
      ||||||:::||||||          :::|||||||||||||   |||
743 SerCysProHisProValGlnAlaLeuAspGlnCysCysProValCysPr 759
```

FIG. 8 (CONT.^{2})

```
582 GGAA.....CCTCACACTCCCTCTGGACTCCGGG 610
      |||           |||     |||:::
759 oGluLysGlnArgSerArgAspLeuProSerLeuProAsn......LeuG 774

611 CCCCACCAAAGTCCTGCCAGCACAACGGGACCATGTACCAACACGGAGAG 660
     |||   ::::::|||     :::|||           ::::::
774 luProGlyGluGlyCysTyrPheAspGlyAspArgSerTrpArgAlaAla 790

661 ATCTTCAGTGCCCATGAGCTGTTCCCCTCCCGCCTGCCCAACCAGTGTGT 710
         |||   :::   |||                 :::|||::
791 GlyThrArgTrpHisProValValProProPheGlyLeuIleLysCysAl 807

711 CCTCTGCAGCTGC......ACAGAGGGCCAGATCTACTGCGGGCTCACAA 754
    ::::|||:::|||     :::   |||::::::::::|||     :::
807 aValCysThrCysLysGlyAlaThrGlyGluValHisCysGluLysValG 824

755 CCTGCCCCGAACCAGGCTGCCCAGCACCCCTCCCGCTG...CCAGACTCC 801
    ||||||:::   :::|||     |||:::         |||   :::
824 lnCysProArgLeuAlaCysAlaGlnProValArgAlaAsnProThrAsp 840
```

FIG. 8 (CONT.[3])

```
802 TGCTGCCAGGCCTGCAAAGATGAGGCAAGTGAGCAATCGGATGAAGAGGA 851
                      ||||||:::   |||
841 CysCysLysGlnCys.............................. 845

852 CCGTGTGCAGTCGCTCCATGGGGTGAGACATCCTCAGGATCCATGTTCCA 901
                         |||   :::|
846 .....................................ProValGlyS 849

902 GTGATGCTGGGAGAAAGAGAGGCCCGGGCACCCCAGCCCCCACTGGCCTC 951
      ||   ::::::   |||   |||                  |||
849 erGlyThrAsnAlaLysLeuGlyAsp.............ProMetGlnAla 861

952 AGCGCCCCTCTGAGCTTCATCCCTCGCCACTTCATACCCAAGGGAGCAGG 1001
                      ::::::|||
862 AspGlyPro..................................... 864

1002 CAGCACAACTGTCAAGATCGTCCTGAAGGAGAAACATAAGAAAGCCTGTG 1051
                      ::::::|||
865 .....................................ArgGlyCysA 868
```

FIG. 8 (CONT.[4])

```
1052 TGCATGGCGGGAAGACGTACTCCCACGGGGAGGTGTGGCACCCGGCCTTC 1101
         :::|||:::   :::   :::::::::   |||||||||:::
 868 rgPheAlaGlyGlnTrpPheProGluAsnGlnSerTrpHisProSerVal 884

1102 CGTGCCTTCGGCCCCTTGCCCTGCATCCTATGCACCTGTGAGGATGGCCG 1151
         ||||||   :::   ||||||   |||   |||       |||
 885 ProProPheGlyGluMetSerCysIleThrCysArgCysGlyAlaGlyVa 901

1152 CCAGGACTGCCAGCGTGTGACCTGTCCCACCGAGTACCCCTGCCGTCACC 1201
           |||:::|||     |||                 |||
 901 lProHisCysGluArgAspAspCysSerProProLeuSerCysGlySerG 918

1202 CCGAGAAAGTGGCTGGGAAGTGCTGCAAGATTTGC..............  1236
             ::::::     ::::::||||||:::   |||
 918 lyLysGlu......SerArgCysCysSerHisCysThrAlaGlnArgSer 932

 1237 ..................CCAGAGGACAAAGCAGACCCTGGCCACAGT 1266
               ||||||   :::   :::      ||||||
  933 SerGluThrArgThrLeuProGluLeuGluLysGluAlaGluHisSer 948
```

FIG. 8 (CONT.[5])

```
862 TCGCTCCATGGGGTGAGACATCCTCAGGATCCATGTTCCAGTGATGCTGG 911
    |||   |||      ::::::          |||      :::   :::||
  3 SerTyrHisArgSerHisTyrAspProProProSerArgGlnAlaGlyGl 19

912 GAGAAAGAGAGGCCCGGGCACCCCAGCCCCCACTGGCCTCAGCGCCCCTC 961
          |   :::|||   ||||||:::   :::       |||
 19 yLeuSerArgPheProGlyAlaArgSerHisArgGlyAlaLeu..... 33

962 TGAGCTTCATCCCTCGCCACTTCAGACCCAAGGGAGCAGGCAGCACAACT 1011
                :::      :::   |||::::::|||
  34 ....................MetAspSerGlnGlnAlaSerGlyThrIle 43

1012 GTCAAGATCGTCCTGAAGGAGAAACATAAG......AAAGCCTGTGTGCA 1055
     |||:::||||||:::::::::||||||||      ::::::||||||
  44 ValGlnIleValIleAsnAsnLysHisLysHisGlyGlnValCysValSe 60

1056 TGGCGGGAAGACGTACTCCCACGGGGAGGTGTGGCACCCGGCCTTCCGTG 1105
     :::||||||||||||||||||||||||   |||||||||   :::||||
  60 rAsnGlyLysThrTyrSerHisGlyGluSerTrpHisProAsnLeuArgA 77
```

FIG. 9

```
1106 CCTTCGGCCCCTTGCCCTGCATCCTATGCACCTGTGAGGATGGCCGCCAG 1155
     ||||||||   :::   |||:::|||||||||||||:::      :::|||
  77 laPheGlyIleValGluCysValLeuCysThrCysAsnValThrLysGln 93

1156 GACTGCCAGCGTGTGACCTGTCCCACCGAGTACCCCTGCCGTCACCCCGA 1205
     :::|||:::::::::   ||||||:::::||||||||||:::::|||::
  94 GluCysLysLysIleHisCysProAsnArgTyrProCysLysTyrProGl 110

1206 GAAAGTGGCTGGGAAGTGCTGCAAGATTTGCCCAGAGGACAAAGCAGAC. 1254
       :|||:::   ||||||||||||||::||||||      ||||||
 110 nLysIleAspGlyLysCysCysLysValCysProGlyLysLysAlaLysG 127

1255 .........CCTGGCCACAGT...GAGATCAGTTCTACCAGGTGTCCCAAG 1293
         ||||||:::|||   :::   ::::::      |||   :::
 127 luGluLeuProGlyGlnSerPheAspAsnLysGlyTyrPheCysGlyGlu 143

1294 GCACCGGGCCGGGTCCTCGTCCACACATCGGTA...TCCCCAAGCCCAGA 1340
           :::   |||:::   ||||||          :::   ::
 144 Glu......ThrMetProValTyrGluSerValPheMetGluAspGlyGl 158
```

FIG. 9 (CONT.[1])

```
1341 CAACCTGCGTCGCTTTGCCCTGGAACACGAGGCCTCGGACTTGGTGGAGA 1390
     ::::   |||::::::|||||||||   |||              ||||||:
 158 uThrThrArgLysIleAlaLeuGluThrGluArgProProGlnValGluV 175

1391 TCTACCTCTGGAAGCTGGTAAANNNNNNNNNNNNNNNNNNNNNNNNNNN 1440
                       ::::::::|||   :::
 175 alHisValTrpThrIle..................................  180

1441 NNNNNNNNNNNNNNNNNNNNNNNNNNNCAGAATCTTCCACTTGACTCAGA 1490
                    |||::: ||||||         |
 181 ............ArgLysGlyIleLeuGlnHis.PheHis.........I 190

1491 TCAAGAAAGTCAGGAAGCAAGACTTCCAGAAAGAGGCACAGCACTTCCGA 1540
     ||:::|||:::   |||:::   |||:::   |||      ||||||:::
 190 leGluLysIleSerLysArgMetPheGlu...GluLeuProHisPheLys 205

1541 CTGCTCGCTGGCCCCCACGAAGGTCACTGGAACGTCTTCCTAGCCCAGAC 1590
          |||::::::              ::::::|||:::::::|||
 206 LeuValThrArgThrThrLeuSerGlnTrpLysIlePheThrGluGlyGl 222
```

FIG. 9 (CONT.[2])

```
1591 CCTGGAGCTGAAGGTCACGGCCAGTCCAGACAAAGTGACCAAGACATAAC 1640
       :::::::::      :::|||                        +++
 222 uAlaGlnIleSerGlnMetCysSer..........................  230

1641 AAAGACCTAACAGTTGCAGATATGAGCTGTATAATTGTTGTTATTATATA 1690
         :::   ||||||   ||||||   :::                :::
 231 .....SerArgValCysArgThrGluLeuGluAspLeuValLysValLeu 245

          1691 TTAATAAATAAGAAGTTGCATAACCAT 1717
               :::::::::::      :::|||
 246 TyrLeuGluArgSerGluLysGlyHis 254
```

FIG. 9 (CONT.[3])

```
  3 CTTCCCCCTTTCTTTGATCGCCTCTCC.............CTTCTGCTGGA 40
    |||||||||      ::::::   :::              |||   |||||
390 LeuProProValLysSerGlnAlaAlaGlyHisAlaTrpLeuSerLeuAs 406

 41 CCTTCCTTCGTCTCTCCATCTCTCCCTCCTT..................... 71
            |              ||||||   :::::::|||
406 pThrHisCysHisLeuHisTyrGluValLeuLeuAlaGlyLeuGlyGlyS 423

 72 ..TCCCCGCGTTCTCTTTCCACCTTTCTCTTCTTCCCACCTTAGACCTCC 119
      :::      ::::::::::::   |||:::   ||||||+++|||
423 erGluGlnGlyThrValThrAlaHisLeuLeuGlyProProGlyThr... 438

120 CTTCCTGCCCTCCTTTCCTGCCCACCGCTGCTTCCTGGCCCTTCTCCGAC 169
                          |||:::|||
439 ..................................ProGlyProAr 442

170 CCCGCTCTAGCAG.............CAGACCTCCTGGGGTCTGTGGGTTG 207
      |||+++:::              :::   :::   |||:::      |
442 gArgLeuLeuLysGlyPheTyrGlySerGluAlaGlnGlyValValLysA 459
```

FIG. 10

```
208 ATCTGTGGCCCCTGTGCCTCCGTGTCCTTTTCGTCTCCCTTCCTCCCGAC 257
                |||||   |||     ||||||   |||
459 spLeuGluProGluLeuLeuArgHisLeu..................... 468

258 TCCGCTCCCGGACCAGCGGCCTGACCCTGGGGAAAGGATGGTTCC...CG 304
     |||   |||   |||:::+++                   ||||||   ||
469 ...AlaLysGlyMetAlaSerLeuLeuIleThrThrLysGlySerProAr 484

305 AGGTGAG.....................................GGTC 315
     |||||||                                     |||
484 gGlyGluLeuArgGlyGlnValHisIleAlaAsnGlnCysGluValGlyG 501

316 CTCTCCTCCTT..............GCTGGGACTCGCGCT......... 341
          |||   |||                 |||:::||||||
501 lyLeuArgLeuGluAlaAlaGlyAlaGluGlyValArgAlaLeuGlyAla 517

342 ..........................................GCTCT 346
                               |||||
518 ProAspProAlaSerAlaAlaProProValValProGlyLeuProAlaLe 534
```

FIG. 10 (CONT.[1])

```
347 GGTTCCC......CCTGGACTCCCACGCTCGAGCCCGCCCAGACATGTTC 390
          |:::|||      ||||||   ||| :::|||   |||
534 uAlaProAlaLysProGlyGlyPro.GlyArgProArgAspProAsnThr 550

391 TGCCTTTTCCATGGGAAGAGATACTCCCCCGGCGAGAGCTGGCACCCCTA 440
     |||:::|||:::|||:::::          |||      |||   |||
551 CysPhePheGluGlyGlnGlnArgProHisGlyAlaArgTrpAlaProAs 567

441 CTTGGAGCCACAAGGCCTGATGTACTGCCTGCGCTGTACCTGCTCAGAGG 490
      :::|||                |||      |||||||||:::::
567 nTyrAspPro.............LeuCysSerLeuCysThrCysGlnArgA 580

491 GCGCCCATGTGAGTTGTTACCGCCTCCACTGTCCGCCTGTCCACTGCCCC 540
     :::   |||   |||      :::   |||||||||      ||||||
580 rgThr...ValIleCysAspProValValCysProProProSerCysPro 595

541 CAGCCTGTGACGGAGCCACAGCAATGCTGTCCCAAGTGTGTGGAA..... 585
      :::||||||      |||:::|||||||||||||   |||   |||
596 HisProValGlnAlaProAspGlnCysCysProValCysProGluLysGl 612
```

FIG. 10 (CONT.[2])

```
586 ..................CCTCACACTCCCTCTGGACTCCGGGCCCCACCAA 619
                 |||       |||   :::   |||   |||   :
612 nAspValArgAspLeuProGlyLeuProArgSer...ArgAspProGlyG 628

620 AGTCCTGCCAGCACAACGGGACCATGTACCAACACGGAGAGATCTTCAGT 669
          :::::|||      :::|||                :::::
628 luGlyCysTyrPheAspGlyAspArgSerTrpArgAlaAlaGlyThrArg 644

670 GCCCATGAGCTGTTCCCCTCCCGCCTGCCCAACCAGTGTGTCCTCTGCAG 719
     |||   :::   |||                    :::|||::::::|||::
645 TrpHisProValValProProPheGlyLeuIleLysCysAlaValCysTh 661

720 CTGCACA......GAGGGCCAGATCTACTGCGGCCTCACAACCTGCCCCG 763
    :|||            |||::::::::::|||     :::   ||||||:
661 rCysLysGlyGlyThrGlyGluValHisCysGluLysValGlnCysProA 678

764 AACCAGGCTGCCCAGCACCCCTCCCGCTG...CCAGACTCCTGCTGCCAG 810
    ::   :::|||      |||:::   :::   |||   :::|||||||:::
678 rgLeuAlaCysAlaGlnProValArgValAsnProThrAspCysCysLys 694
```

FIG. 10 (CONT.[3])

```
811 GCCTGCAAA...GGTGAGGCAAGTGAGCAATCGGATGAAGAGGACAGTGT 857
                  |||      |||:::::::::
695 GlnCysProValGlySerGlyAla.............................. 702

858 GCAGTCGCTCCATGGGGTGAGACATCCTCAG......GATCCATGTTCCA 901
              |||||||||      ||||||   ::::
703 ......................HisProGlnLeuGlyAspProMetGlnA 712

902 GTGATGCTGGGAGAAAGAGAGGCCCGGGCACCCCAGCCCCCACTGGCCTC 951
                ::|||                 ||||||
712 laAsp..............GlyPro.......................... 715

952 AGCGCCCCTCTGAGCTTCATCCCTCGCCACTTCAGACCCAAGGGAGCAGG 1001

715 ..................................................... 715

1002 CAGCACAACTGTCAAGATCGTCCTGAAGGAGAAACATAAGAAAGCCTGTG 1051
                              ::::::|||
716 ...................................ArgGlyCysA 719
```

FIG. 10 (CONT.[4])

```
1052 TGCATGGCGGGAAGACGTACTCCCACGGGGAGGTGTGGCACCCGGCCTTC 1101
         :::|||:::   :::   :::::::::   |||||||||:::
 719 rgPheAlaGlyGlnTrpPheProGluSerGlnSerTrpHisProSerVal 735

1102 CGTGCCTTCGGCCCCTTGCCCTGCATCCTATGCACCTGTGAGGATGGCCG 1151
         ||||||   :::   ||||||   |||   |||       |||
 736 ProProPheGlyGluMetSerCysIleThrCysArgCysGlyAlaGlyVa 752

1152 CCAGGACTGCCAGCGTGTGACCTGTCCCACCGAGTACCCCTGCCGTCACC 1201
          |||:::|||      |||                  |||
 752 lProHisCysGluArgAspAspCysSerLeuProLeuSerCysGlySerG 769

1202 CCGAGAAAGTGGCTGGGAAGTGCTGCAAGATTTGC..............  1236
           ::::::     ::::::||||||:::   |||
 769 lyLysGlu......SerArgCysCysSerArgCysThrAlaHisArgArg 783

1237 ......CCAGAGGACAAAGCAGACCCT 1257
           ||||||   ::::::||||||
 784 ProAlaProGluThrArgThrAspPro 792
```

FIG. 10 (CONT.[5])

```
655 TCGCTCCATGGGGTGAGACATCCTCAGGATCCATGTTCCAGTGATGCTGG 704
    |||   |||      ::::::          |||      :::   :::||
  3 SerTyrHisArgSerHisTyrAspProProProSerArgGlnAlaGlyGl 19

705 GAGAAAGAGAGGCCCGGGCACCCCAGCCCCCACTGGCCTCAGCGCCCCTC 754
          |   :::|||   ||||||:::   :::      |||
 19 yLeuSerArgPheProGlyAlaArgSerHisArgGlyAlaLeu..... 33

755 TGAGCTTCATCCCTCGCCACTTCAGACCCAAGGGAGCAGGCAGCACAACT 804
               :::      :::   |||:::::::|||
 34 .....MetAspSerGlnGlnAlaSerGlyThrIle 43

805 GTCAAGATCGTCCTGAAGGAGAAACATAAG......AAAGCCTGTGTGCA 848
     |||:::||||||:::::::::||||||||      ::::::||||||
 44 ValGlnIleValIleAsnAsnLysHisLysHisGlyGlnValCysValSe 60

849 TGGCGGGAAGACGTACTCCCACGGGGAGGTGTGGCACCCGGCCTTCCGTG 898
    :::|||||||||||||||||||||||   |||||||||   :::||||
 60 rAsnGlyLysThrTyrSerHisGlyGluSerTrpHisProAsnLeuArgA 77
```

FIG. 11

```
899 CCTTCGGCCCCTTGCCCTGCATCCTATGCACCTGTGAGGATGGCCGCCAG 948
    ||||||||   :::   |||:::||||||||||||:::      :::|||
 77 laPheGlyIleValGluCysValLeuCysThrCysAsnValThrLysGln 93

949 GACTGCCAGCGTGTGACCTGTCCCACCGAGTACCCCTGCCGTCACCCCGA 998
    :::|||:::::::::   ||||||::::::||||||||::::::|||::
 94 GluCysLysLysIleHisCysProAsnArgTyrProCysLysTyrProGl 110

999 GAAAGTGGCTGGGAAGTGCTGCAAGATTTGCCCAGAGGACAAAGCAGAC. 1047
      :|||:::   |||||||||||||||::||||||      ||||||
110 nLysIleAspGlyLysCysCysLysValCysProGlyLysLysAlaLysG 127

1048 .........CCTGGCCACAGT...GAGATCAGTTCTACCAGGTGTCCCAAG 1086
         ||||||:::|||   :::   ::::::      |||   :::
 127 luGluLeuProGlyGlnSerPheAspAsnLysGlyTyrPheCysGlyGlu 143

1087 GCACCGGGCCGGGTCCTCGTCCACACATCGGTA...TCCCCAAGCCCAGA 1133
           :::   |||:::   ||||||          :::   ::
 144 Glu......ThrMetProValTyrGluSerValPheMetGluAspGlyGl 158
```

FIG. 11 (CONT.[1])

```
1134 CAACCTGCGTCGCTTTGCCCTGGAACACGAGGCCTCGGACTTGGTGGAGA 1183
     ::::   |||::::::|||||||||   |||              ||||||:
 158 uThrThrArgLysIleAlaLeuGluThrGluArgProProGlnValGluV 175

1184 TCTACCTCTGGAAGCTGGTAAANNNNNNNNNNNNNNNNNNNNNNNNNNN 1233
                     ::::::::|||   :::
 175 alHisValTrpThrIle..............................  180

1234 NNNNNNNNNNNNNNNNNNNNNNNNNNNCAGAATTTCCACTTGACTCAGAT 1283
                  |||:::||||||         ||
 181 ............ArgLysGlyIleLeuGlnHisPheHis.........Il 190

1284 CAAGAAAGTCAGGAAGCAAGACTTCCAGAAAGAGGCACAGCACTTCCGAC 1333
     |:::|||:::   |||:::   |||:::   |||      ||||||:::|
 190 eGluLysIleSerLysArgMetPheGlu...GluLeuProHisPheLysL 206
```

FIG. 11 (CONT.[2])

```
1334 TGCTCGCTGGCCCCCACGAAGGTCACTGGAACGTCTTCCTAGCCCAGACC 1383
          ||:::::::                    ::::::|||:::::::|||
 206 euValThrArgThrThrLeuSerGlnTrpLysIlePheThrGluGlyGlu 222

1384 CTGGAGCTGAAGGTCACGGCCAGTCCAGACAAAGTGACCAAGACATAACA 1433
                                                    +++
       :::::::::       :::|||
 223 AlaGlnIleSerGlnMetCysSer..........................  230

1434 AAGACCTAACAGTTGCAGATATGAGCTGTATAATTGTTGTTATTATATAT 1483
           :::    ||||||    ||||||    :::                :::
 231 ....SerArgValCysArgThrGluLeuGluAspLeuValLysValLeuT 246

1484 TAATAAATAAGAAGTTGCATAACCAT 1509
       ::::::::::::       :::|||
 246 yrLeuGluArgSerGluLysGlyHis 254
```

FIG. 11 (CONT.[3])

```
 80 GAAGCTGAGGCTGCCGGCTTTACCTTCTTGCTCTGGTTGGGGCTG..... 124
    :::::::::|||||||||   :::          ||||||:::|||
394 LysSerGlnAlaAlaGlyHisAla.........TrpLeuSerLeuAspTh 407

125 ...................GCCCAGACACTGAGGGGCTGGAGGGCTGTG 155
                       |||   ||||||   |||
407 rHisCysHisLeuHisTyrGluValLeuLeuAlaGlyLeuGlyGlySerG 424

156 GTAGAGGTCATGGGAGGGAGGGACTCAGTCAGATGTAGGTATCAGAGGGA 205
     :::|||                |||          +++   :::   |||
424 luGlnGlyThrValThrAlaHisLeuLeuGlyProProGlyThrProGly 440

206 CCTCTTAGTTAGCTGATAAGGGGAATGGCTGGCAAGGCCCAGGCCAGAGC 255
    |||      +++|||::::::|||:::   |||   ::::::|||:::::
441 ProArgArg...LeuLeuLysGlyPheTyrGly...SerGluAlaGlnGl 455
```

FIG. 12

```
256 TTGGTTTAAATATCAAGCTGGGGTCAAATGCAAAATCATCAGAAAAGTGG 305
    :      |||              ::::::   :::::::::|||   :::|
455 yValValLys.........AspLeuGluProGluLeuLeuArgHisLeuA 469

306 CC..............TTGTTAATTTCAGCAAAG..............  325
          ||                 ||||||||::::::|||
469 laLysGlyMetAlaSerLeuLeuIleThrThrLysGlySerProArgGly 485

326 ..............ATTCACATG.........................  334
                        :::|||:::
486 GluLeuArgGlyGlnValHisIleAlaAsnGlnCysGluValGlyGlyLe 502

335 ..........................................A 335

502 uArgLeuGluAlaAlaGlyAlaGluGlyValArgAlaLeuGlyAlaProA 519

336 AACCTCATTTTCTTCTTCCTCCTGCCCCTCCC......CCACTGCA.GAA 378
        |||              ||||||:::   |||      |||
519 spProAlaSerAlaAlaProProValValProGlyLeuProAlaLeuAla 535
```

FIG. 12 (CONT.[1])

```
379 CCTCACACTCCCTCTGGACTCCGGGCCCCA......CCAAAGTCCTGCCA 422
     |||      |||:::|||          |||      |||:::::::|||
536 ProAlaLysProGlyGlyProGlyArgProArgAspProAsnThrCysPh 552

423 GCACAACGGGACCATGTACCAACACGGAGAGATCTTCAGTGCCCATGAGC 472
          :::|||   :::      ||||||      ::::::
552 ePheGluGlyGlnGlnArgProHisGlyAlaArgTrpAla..... 565

473 TGTTCCCCTCCCGCCTGCCCAACCAG.....TGTGTCCTCTGCAGC 513
           ||||||            |||   ||||||:::
566 ................ProAsnTyrAspProLeuCysSerLeuCysThr 576

514 TGCACAGAGGGCCAGATCTACTGCGGCCTCACAACCTGCCCCGAACCAGG 563
    |||   :::      :::   |||      ::::::||||||   |||::
577 CysGlnArgArgThrValIleCysAspProValValCysProProProSe 593

564 CTGCCCAGCACCCCTCCCGCTGCCAGACTCCTGCTGCCAGGCCTGCAAAG 613
      :||||||   |||:::      ||||||:::||||||   :::|||
593 rCysProHisProValGlnAlaProAspGlnCysCysProValCys.... 608
```

FIG. 12 (CONT.[2])

```
614 ATGAGGCAAGTGAGCAATCGGATGAAGAGGACAGTGTGCAGTCGCTCCAT 663
                |||:::::::::|||:::::|||
609 ....................ProGluLysGlnAspValArgAspLeuPro 618

664 GGGGTGAGACATCCTCAGGATCCA.........TGTTCCAGTGATGCTGG 704
     |||:::   :::   :::||||||         |||     |||:::
619 GlyLeuProArgSerArgAspProGlyGluGlyCysTyrPheAspGlyAs 635

705 GAGAAAG...AGAGGCCCGGGCACC..............CCAGCCCCCA 736
     |||:::   |||:::   ||||||                  |||
635 pArgSerTrpArgAlaAlaGlyThrArgTrpHisProValValProProP 652

737 CTGGCCTC..................................... 744
                          ||||||
652 heGlyLeuIleLysCysAlaValCysThrCysLysGlyGlyThrGlyGlu 668

745 ...................................AGCGCCCCTCT 755
                          :::   |||:
669 ValHisCysGluLysValGlnCysProArgLeuAlaCysAlaGlnProVa 685
```

FIG. 12 (CONT.[3])

```
756 GAGCTTCATCCCTCGCCACTTC.........AGACCCAAGGGAGCAGGCA 796
    :         |||                       |||   |||:::|||:
685 lArgValAsnProThrAspCysCysLysGlnCysProValGlySerGlyA 702

797 GC...ACAACTGTCAAGATCGTCCTGAAGGAGAAACATAAGAAAGCCTGT 843
    ::          :::         ::::::                :::::|||
702 laHisProGlnLeuGlyAspProMetGlnAlaAspGlyProArgGlyCys 718

844 GTGCATGGCGGGAAGACGTACTCCCACGGGGAGGTGTGGCACCCGGCCTT 893
        :::|||:::   :::   :::::::::   |||||||||:::
719 ArgPheAlaGlyGlnTrpPheProGluSerGlnSerTrpHisProSerVa 735

894 CCGTGCCTTCGGCCCCTTGCCCTGCATCCTATGCACCTGTGAGGATGGCC 943
        ||||||   :::   ||||||   |||   |||       |||
735 lProProPheGlyGluMetSerCysIleThrCysArgCysGlyAlaGlyV 752

944 GCCAGGACTGCCAGCGTGTGACCTGTCCCACCGAGTACCCCTGCCGTCAC 993
          |||:::|||      |||                  |||
752 alProHisCysGluArgAspAspCysSerLeuProLeuSerCysGlySer 768
```

FIG. 12 (CONT.[4])

```
994 CCCGAGAAAGTGGCTGGGAAGTGCTGCAAGATTTGC.............. 1029
              ::::::      ::::::||||||:::   |||
769 GlyLysGlu......SerArgCysCysSerArgCysThrAlaHisArgAr 783

          1030 .......CCAGAGGACAAAGCAGACCCT 1050
                     ||||||   ::::::||||||
           783 gProAlaProGluThrArgThrAspPro 792
```

FIG. 12 (CONT.[5])

```
486 TGCCTTTTCCATGGGAAGAGATACTCCCCCGGCGAGAGCTGGCACCCCTA 535
    |||:::|||:::|||:::::::::::::   |||:::::::|||   |||
691 CysPhePheGluGlyGluGlnHisThrHisGlySerGlnTrpThrProGl 707

536 CTTGGAGCCACAAGGCCTGATGTACTGCCTGCGCTGTACCTGCTCAGAGG 585
      :::                    |||:::   |||||||||:::::::
707 nTyrAsnThr..............CysPheThrCysThrCysGlnLysL 719

586 GCGCCCATGTGAGTTGTTACCGCCTCCACTGTCCGCCTGTCCACTGCCCC 635
      :::   |||   |||       :::   ||||||   :::   |||
719 ysThr...ValIleCysAspProValMetCysProThrLeuSerCysThr 734

636 CAGCCTGTGACGGAGCCACAGCAATGCTGTCCCAAGTGTGTGGAACCTCA 685
      :::   |||         :::||||||||||||   |||   |||
735 HisThrValGlnProGluAspGlnCysCysProIleCysGluGluLysLy 751

686 CACTCCCTCTGGACTCCGGGCC.................CCACCAAAGT 717
      :::         |||                     |||::::
751 sGluSerLysGluThrAlaAlaValGluLysValGluGluAsnProGluG 768
```

FIG. 13

```
718 CCTGCCAGCACAACGGGACCATGTACCAACACGGAGAGATCTTCAGTGCC 767
      ::|||       :::|||   :::   :::|||:::          :::
768 lyCysTyrPheGluGlyAspGlnLysMetHisAlaProGlyThrThrTrp 784

768 CATGAGCTGTTCCCCTCCCGCCTGCCCAACCAGTGTGTCCTCTGCAGCTG 817
    |||   :::   |||                :::|||::::::|||:::||
785 HisProPheValProProPheGlyTyrIleLysCysAlaValCysThrCy 801

818 C......ACAGAGGGCCAGATCTACTGCGGCCTCACAACCTGCCCCGAAC 861
      |      :::   |||::::::::::|||      :::||||||||||
801 sLysGlySerThrGlyGluValHisCysGluLysValThrCysProProL 818

862 CAGGCTGCCCAGCACCCCTCCCGCTG...CCAGACTCCTGCTGCCAAGCC 908
        |||     |||:::          |||::::::|||||||:::
818 euThrCysSerArgProIleArgArgAsnProSerAspCysCysLysGlu 834

909 TGCAAAGATGAGGCAAGTGAGCAATCGGATGAAGAGGACAGTGTGCAGTC 958
    |||     |||   :::          :::::::|||:::   :::|||::
835 CysProProGluGluThrProProLeuGluAspGluGluMetMetGlnAl 851
```

FIG. 13 (CONT.[1])

```
959 GCTCCATGGGGTGAGACATCCTCAGGATCCATGTTCCAGTGATGCTGGGA 1008
                                    :
851 a..................................................  851

1009 GAAAGAGAGGCCCGGGCACCCCAGCCCCCACTGGCCTCAGCGCCCCTCTG 1058
                                    ||||||
852 ............AspGlyThr.............................  854

1059 AGCTTCATCCCTCGCCACTTCAGACCCAAGGGAGCAGGCAGCACAACTGT 1108
854 ..................................................  854

1109 CAAGATCGTCCTGAAGGAGAAACATAAGAAAGCCTGTGTGCATGGCGGGA 1158
                    :::   |||      |||   :
855 ............................ArgLeuCysLysPheGlyLysA 862

1159 AGACGTACTCCCACGGGGAGGTGTGGCACCCGGCCTTCCGTGCCTTCGGC 1208
     ::   |||::::::::::|||   |||||||.||:::          |||
862 snTyrTyrGlnAsnSerGluHisTrpHisProSerValProLeuValGly 878
```

FIG. 13 (CONT.[2])

```
1209 CCCTTGCCCTGCATCCTATGCACCTGTGAGGATGGCCGCCAGGACTGCCA 1258
      :::   ||||||   |||   |||:::   |||          |||||
 879 GluMetLysCysIleThrCysTrpCysAspHisGlyValThrLysCysGl 895

1259 GCGTGTGACCTGTCCCACCGAGTACCCCTGCCGTCACCCCGAGAAAGTGG 1308
      ||||     ||||||           ||||||:::|||   ::::::
 895 nArgLysGlnCysProLeu...LeuSerCysArgAsnProIleArgThrG 911

          1309 CTGGGAAGTGCTGCAAGATTTGCCCAGAGGAC 1340
               ||||||||||||      |||   ||||||
           911 luGlyLysCysCysProGluCysIleGluAsp 921
```

FIG. 13 (CONT.[3])

```
462 GGTTTGCCAGGCCCAGACATGTTCTGCCTTTTCCATGGGAAGAGATACTC 511
      |||   |||                    |||:::|||:::|||::::::
543 GlyArgProArgAspProAsnThrCysPhePheGluGlyGlnGlnArgPr 559

512 CCCCGGCGAGAGCTGGCACCCCTACTTGGAGCCACAAGGCCTGATGTACT 561
      |||      |||   |||      :::|||                   |
559 oHisGlyAlaArgTrpAlaProAsnTyrAspPro.............LeuC 572

562 GCCTGCGCTGTACCTGCTCAGAGGGCGCCCATGTGAGTTGTTACCGCCTC 611
    ||      |||||||||:::::   :::   |||   |||       :::
572 ysSerLeuCysThrCysGlnArgArgThr...ValIleCysAspProVal 587

612 CACTGTCCGCCTGTCCACTGCCCCCAGCCTGTGACGGAGCCACAGCAATG 661
      |||||||||     ||||||:::||||||      |||:::|||||
588 ValCysProProProSerCysProHisProValGlnAlaProAspGlnCy 604

662 CTGTCCCAAGTGTGTGGAA.....................CCTCACACTC 690
    |||||||   |||   |||                    |||      |
604 sCysProValCysProGluLysGlnAspValArgAspLeuProGlyLeuP 621
```

FIG. 14

```
691 CCTCTGGACTCCGGGCCCCACCAAAGTCCTGCCAGCACAACGGGACCATG 740
       ||   :::   |||   |||   :::::::|||      :::|||
621 roArgSer...ArgAspProGlyGluGlyCysTyrPheAspGlyAspArg 636

741 TACCAACACGGAGAGATCTTCAGTGCCCATGAGCTGTTCCCCTCCCGCCT 790
          ::::::                  |||   :::   |||
637 SerTrpArgAlaAlaGlyThrArgTrpHisProValValProProPheGl 653

791 GCCCAACCAGTGTGTCCTCTGCAGCTGCACA......GAGGGCCAGATCT 834
       :::|||::::::|||:::|||                |||:::::::
653 yLeuIleLysCysAlaValCysThrCysLysGlyGlyThrGlyGluValH 670

835 ACTGCGGCCTCACAACCTGCCCCGAACCAGGCTGCCCAGCACCCCTCCCG 884
     ::|||      :::   ||||||:::   :::|||       |||:::
670 isCysGluLysValGlnCysProArgLeuAlaCysAlaGlnProValArg 686

885 CTG...CCAGACTCCTGCTGCCAAGCCTGCAAAGATGAGGCAAGTGAGCA 931
            :::   |||   :::|||||:|:::   |||
687 ValAsnProThrAspCysCysLysGlnCys..................... 696
```

FIG. 14 (CONT.[1])

```
932 ATCGGATGAAGAGGACAGTGTGCAGTCGCTCCATGGGGTGAGACATCCTC 981
                    :::   :::      |||||||
697 ........................ProValGlySerGlyAlaHisProG 705

982 AG......GATCCATGTTCCAGTGATGCTGGGAGAAAGAGAGGCCCGGGC 1025
      ||      ||||||   :::::::|||                  ||||||
705 lnLeuGlyAspProMetGlnAlaAsp.................GlyPro... 715

1026 ACCCCAGCCCCCACTGGCCTCAGCGCCCCTCTGAGCTTCATCCCTCGCCA 1075
 715 ..................................................... 715

1076 CTTCAGACCCAAGGGAGCAGGCAGCACAACTGTCAAGATCGTCCTGAAGG 1125
 715 ..................................................... 715

1126 AGAAACATAAGAAAGCCTGTGTGCATGGCGGGAAGACGTACTCCCACGGG 1175
           ::::::|||      :::|||:::   :::   ::::::
 716 ...........ArgGlyCysArgPheAlaGlyGlnTrpPheProGluSer 728
```

FIG. 14 (CONT.[2])

```
1176 GAGGTGTGGCACCCGGCCTTCCGTGCCTTCGGCCCCTTGCCCTGCATCCT 1225
      :::   |||||||||:::          ||||||   :::   ||||||
 729 GlnSerTrpHisProSerValProProPheGlyGluMetSerCysIleTh 745

1226 ATGCACCTGTGAGGATGGCCGCCAGGACTGCCAGCGTGTGACCTGTCCCA 1275
       |||   |||      |||          |||:::|||      |||
 745 rCysArgCysGlyAlaGlyValProHisCysGluArgAspAspCysSerL 762

1276 CCGAGTACCCCTGCCGTCACCCCGAGAAAGTGGCTGGGAAGTGCTGCAAG 1325
           |||          ::::::      :::::|||||||:::
 762 euProLeuSerCysGlySerGlyLysGlu......SerArgCysCysSer 776

1326 ATTTGC.....................CCAGAGGACAAAGCAGACCCT 1352
       |||                       ||||||   :::::|||||||
 777 ArgCysThrAlaHisArgArgProAlaProGluThrArgThrAspPro 792
```

FIG. 14 (CONT.[3])

```
243 GAGACAGTGGCATGCCCAGTGTTGCACAGTAAGTGTGTGGTAAAGCCGAG 292
    :::   |||   ||||||:::|||              :::|||:::|||::
724 AspProValMetCysProThrLeuSerCysThrHisThrValGlnProGl 740

293 ATTCAAACTCAGACCTTCTGGCCCCTTGCCTAGGAGAGCATGCCCAGTTG 342
    :                                    ::::::||||||:::
740 u.............................AspGlnCysCysProIle. 746

343 TCTAGCAGATTCTCTTTTGCCTGAGTGGCCCAGATGACATCTCTTTTAGA 392
        +++              |||   :::      ::::::   :::
747 ..................CysGluGluLysLysGluSerLysGluThrAla 757

393 GCTAGAAAGAAGGAGAAATGAGACAGGGTCTTTGGGCTGGAGCCTCCTGG 442
    |||   :::|||   :::+++                    :::|||   ||
758 AlaValGluLysValGlu...................GluAsnProGluGl 768

443 GACTAACATGGCACTGGTCGGTTTGCCAGGCCCAGACATGTTCTGCCTTT 492
    |                                              |||   |
768 y.....................................CysTyrP 771

493 TCCATGGG......AAGAGATACTCCCCCGGCGAGAGCTGGCACCCCTAC 536
    ||:::|||      |||   ::::::||||||   :::||||||||||:::
771 heGluGlyAspGlnLysMetHisAlaProGlyThrThrTrpHisProPhe 787
```

FIG. 15

```
537 TTGGAGCCACAAGGCCTGATGTACTGCCTGCGCTGTACCTGC...TCAGA 583
    :::   |||   |||   :::   |||       |||||||||   :::::
788 ValProProPheGlyTyrIleLysCysAlaValCysThrCysLysGlySe 804

584 GGGCGCCCATGTGAGTTGTTACCGCCTCCACTGTCCGCCTGTCCACTGCC 633
     :   ::::::|||   |||   :::::   |||||||||:::   |||
804 rThrGlyGluValHisCysGluLysValThrCysProProLeuThrCysS 821

634 CCCAGCCTGTG...ACGGAGCCACAGCAATGCTGTCCCAAGTGTGTGGAA 680
        :::|||:::      :::|||:::::::||||||   :::|||
821 erArgProIleArgArgAsnProSerAspCysCysLysGluCysProPro 837

681 CCTCACACTCCC.....................................TCTGGACT 700
      :::||||||                                      :::|||
838 GluGluThrProProLeuGluAspGluGluMetMetGlnAlaAspGlyTh 854

701 CCGGGCCCCACCAAAGTCCTGCCAGCACAACGGGACCATGTACCAACACG 750
     |||                |||:::   :::   :::   ||||||::::
854 rArgLeu.............CysLysPheGlyLysAsnTyrTyrGlnAsnS 867

751 GAGAGATCTTCAGTGCCCATGAGCTGTTCCCCTCCCGCCTGCCCAACCAG 800
    ::|||   :::      |||         |||                :::
867 erGluHisTrp......HisProSerValProLeuValGlyGluMetLys 881
```

FIG. 15 (CONT.[1])

```
801 TGTGTCCTCTGCAGCTGCACAGAGGGCCAGATCTACTGCGGCCTCACAAC 850
        |||:::   |||   |||   :::|||          |||
882 CysIleThrCysTrpCysAspHisGlyValThrLysCysGlnArgLysGl 898

851 CTGCCCCGAACCAGGCTGCCCAGCACCCCTCCCGCTGCCAGACTCCTGCT 900
    ||||||     :::|||     |||:::                :::||||
898 nCysProLeuLeuSerCysArgAsnProIleArgThrGluGlyLysCysC 915

901 GCCAAGCCTGCAAAGATGAGGCAAGTGAGCAATCGGATGAAGAGGACAGT 950
        ||     |||   ::::::     |||::::::::::
915 ysProGluCysIleGluAspPheMetGluLysGluGluMetAlaLysMet 931

951 GTGCAGTCGCTCCATGGGGTGAGACAT 977
    :::::::::      :::   ||||||
932 AlaGluLysLysLysSerTrpArgHis 940
```

FIG. 15 (CONT.^{2})

```
462 GGTTTGCCAGGCCCAGACATGTTCTGCCTTTTCCATGGGAAGAGATACTC 511
      |||   |||                    |||:::|||:::|||::::::
543 GlyArgProArgAspProAsnThrCysPhePheGluGlyGlnGlnArgPr 559

512 CCCCGGCGAGAGCTGGCACCCCTACTTGGAGCCACAAGGCCTGATGTACT 561
      |||      |||   |||      :::|||                   |
559 oHisGlyAlaArgTrpAlaProAsnTyrAspPro.............LeuC 572

562 GCCTGCGCTGTACCTGCTCAGAGGGCGCCCATGTGAGTTGTTACCGCCTC 611
    ||      |||||||||:::::   :::   |||   |||      :::
572 ysSerLeuCysThrCysGlnArgArgThr...ValIleCysAspProVal 587

612 CACTGTCCGCCTGTCCACTGCCCCCAGCCTGTGACGGAGCCACAGCAATG 661
      |||||||||      ||||||:::||||||      |||:::|||||
588 ValCysProProProSerCysProHisProValGlnAlaProAspGlnCy 604

662 CTGTCCCAAGTGTGTGGAA.....................CCTCACACTC 690
    |||||||   |||   |||                    |||      |
604 sCysProValCysProGluLysGlnAspValArgAspLeuProGlyLeuP 621
```

FIG. 16

```
691 CCTCTGGACTCCGGGCCCCACCAAAGTCCTGCCAGCACAACGGGACCATG 740
      ||   :::   |||   |||   :::::::|||       :::|||
621 roArgSer...ArgAspProGlyGluGlyCysTyrPheAspGlyAspArg 636

741 TACCAACACGGAGAGATCTTCAGTGCCCATGAGCTGTTCCCCTCCCGCCT 790
        ::::::                    |||   :::   |||
637 SerTrpArgAlaAlaGlyThrArgTrpHisProValValProProPheGl 653

791 GCCCAACCAGTGTGTCCTCTGCAGCTGCACA......GAGGGCCAGATCT 834
      :::|||::::::|||:::|||                |||:::::::
653 yLeuIleLysCysAlaValCysThrCysLysGlyGlyThrGlyGluValH 670

835 ACTGCGGCCTCACAACCTGCCCCGAACCAGGCTGCCCAGCACCCCTCCCG 884
     ::|||      :::   ||||||:::   :::|||       |||:::
670 isCysGluLysValGlnCysProArgLeuAlaCysAlaGlnProValArg 686

885 CTG...CCAGACTCCTGCTGCCAAGCCTGCAAAGATGAGGCAAGTGAGCA 931
            :::   |||   :::|||||-|:::   |||
687 ValAsnProThrAspCysCysLysGlnCys..................... 696
```

FIG. 16 (CONT.[1])

```
932 ATCGGATGAAGAGGACAGTGTGCAGTCGCTCCATGGGGTGAGACATCCTC 981
                :::   :::      |||||||
697 ........................ProValGlySerGlyAlaHisProG 705

      982 AG......GATCCATGTTCCAGTGATGCTGGGAGA 1010
          ||      ||||||   :::::::|||:::   |||
      705 lnLeuGlyAspProMetGlnAlaAspGlyProArg 716
```

FIG. 16 (CONT.[2])

```
243 GAGACAGTGGCATGCCCAGTGTTGCACAGTAAGTGTGTGGTAAAGCCGAG 292
    :::   |||   ||||||:::|||              :::|||:::|||::
724 AspProValMetCysProThrLeuSerCysThrHisThrValGlnProGl 740

293 ATTCAAACTCAGACCTTCTGGCCCCTTGCCTAGGAGAGCATGCCCAGTTG 342
    :                                    ::::::||||||:::
740 u.........................AspGlnCysCysProIle. 746

343 TCTAGCAGATTCTCTTTTGCCTGAGTGGCCCAGATGACATCTCTTTTAGA 392
        +++             |||   :::      ::::::   :::
747 .................CysGluGluLysLysGluSerLysGluThrAla 757

393 GCTAGAAAGAAGGAGAAATGAGACAGGGTCTTTGGGCTGGAGCCTCCTGG 442
    |||   :::|||   :::+++                    :::|||   ||
758 AlaValGluLysValGlu.................GluAsnProGluGl 768

443 GACTAACATGGCACTGGTCGGTTTGCCAGGCCCAGACATGTTCTGCCTTT 492
    |                                              |||   |
768 y.....................................CysTyrP 771
```

FIG. 17

```
493 TCCATGGG......AAGAGATACTCCCCCGGCGAGAGCTGGCACCCCTAC 536
    ||:::|||       |||   ::::::||||||   :::||||||||||:::
771 heGluGlyAspGlnLysMetHisAlaProGlyThrThrTrpHisProPhe 787

537 TTGGAGCCACAAGGCCTGATGTACTGCCTGCGCTGTACCTGC...TCAGA 583
    :::   |||   |||   :::   |||       |||||||||   :::::
788 ValProProPheGlyTyrIleLysCysAlaValCysThrCysLysGlySe 804

584 GGGCGCCCATGTGAGTTGTTACCGCCTCCACTGTCCGCCTGTCCACTGCC 633
     :   ::::::|||.   |||   ::::::   |||||||||:::   |||
804 rThrGlyGluValHisCysGluLysValThrCysProProLeuThrCysS 821

634 CCCAGCCTGTG...ACGGAGCCACAGCAATGCTGTCCCAAGTGTGTGGAA 680
        :::|||:::      :::|||:::::::||||||   :::|||
821 erArgProIleArgArgAsnProSerAspCysCysLysGluCysProPro 837

681 CCTCACACTCCC.....................TCTGGACT 700
      :::||||||                     :::|||
838 GluGluThrProProLeuGluAspGluGluMetMetGlnAlaAspGlyTh 854
```

FIG. 17 (CONT.[1])

```
701 CCGGGCCCCACCAAAGTCCTGCCAGCACAACGGGACCATGTACCAACACG 750
    |||               |||:::   :::   :::   ||||||::::
854 rArgLeu............CysLysPheGlyLysAsnTyrTyrGlnAsnS 867

751 GAGAGATCTTCAGTGCCCATGAGCTGTTCCCCTCCCGCCTGCCCAACCAG 800
    ::|||   :::      |||         |||                :::
867 erGluHisTrp......HisProSerValProLeuValGlyGluMetLys 881

801 TGTGTCCTCTGCAGCTGCACAGAGGGCCAGATCTACTGCGGCCTCACAAC 850
        |||:::   |||   |||   :::|||          |||
882 CysIleThrCysTrpCysAspHisGlyValThrLysCysGlnArgLysGl 898

851 CTGCCCCGAACCAGGCTGCCCAGCACCCCTCCCGCTGCCAGACTCCTGCT 900
    ||||||      :::|||      |||:::              :::||||
898 nCysProLeuLeuSerCysArgAsnProIleArgThrGluGlyLysCysC 915

901 GCCAAGCCTGCAAAGATGAGGCAAGTGAGCAATCGGATGAAGAGGACAGT 950
        ||      |||   ::::::      |||:::::::::
915 ysProGluCysIleGluAspPheMetGluLysGluGluMetAlaLysMet 931

951 GTGCAGTCGCTCCATGGGGTGAGACAT 977
    :::::::::      :::   ||||||
932 AlaGluLysLysLysSerTrpArgHis 940
```

FIG. 17 (CONT.^{2})

```
462 GGTTTGCCAGGCCCAGACATGTTCTGCCTTTTCCATGGGAAGAGATACTC 511
      |||   |||                  |||:::|||:::|||::::::
543 GlyArgProArgAspProAsnThrCysPhePheGluGlyGlnGlnArgPr 559

512 CCCCGGCGAGAGCTGGCACCCCTACTTGGAGCCACAAGGCCTGATGTACT 561
      |||      |||   |||      :::|||                  |
559 oHisGlyAlaArgTrpAlaProAsnTyrAspPro.............LeuC 572

562 GCCTGCGCTGTACCTGCTCAGAGGGCGCCCATGTGAGTTGTTACCGCCTC 611
    ||      |||||||||:::::   :::   |||   |||      :::
572 ysSerLeuCysThrCysGlnArgArgThr...ValIleCysAspProVal 587

612 CACTGTCCGCCTGTCCACTGCCCCCAGCCTGTGACGGAGCCACAGCAATG 661
      |||||||||      ||||||:::|||||      |||:::|||||
588 ValCysProProProSerCysProHisProValGlnAlaProAspGlnCy 604

662 CTGTCCCAAGTGTGTGGAA.....................CCTCACACTC 690
    |||||||   |||   |||                    |||      |
604 sCysProValCysProGluLysGlnAspValArgAspLeuProGlyLeuP 621
```

FIG. 18

```
691 CCTCTGGACTCCGGGCCCCACCAAAGTCCTGCCAGCACAACGGGACCATG 740
       ||    :::    |||    |||    :::::::|||       :::|||
621 roArgSer...ArgAspProGlyGluGlyCysTyrPheAspGlyAspArg 636

741 TACCAACACGGAGAGATCTTCAGTGCCCATGAGCTGTTCCCCTCCCGCCT 790
          ::::::                    |||    :::    |||
637 SerTrpArgAlaAlaGlyThrArgTrpHisProValValProProPheGl 653

791 GCCCAACCAGTGTGTCCTCTGCAGCTGCACA......GAGGGCCAGATCT 834
       :::|||::::::|||::::|||                |||::::::
653 yLeuIleLysCysAlaValCysThrCysLysGlyGlyThrGlyGluValH 670

835 ACTGCGGCCTCACAACCTGCCCCGAACCAGGCTGCCCAGCACCCCTCCCG 884
      ::|||       :::    ||||||:::    :::|||       |||:::
670 isCysGluLysValGlnCysProArgLeuAlaCysAlaGlnProValArg 686

885 CTG...CCAGACTCCTGCTGCCAAGCCTGCAAAGATGAGGCAAGTGAGCA 931
              :::    |||    :::|||||:|:::    |||
687 ValAsnProThrAspCysCysLysGlnCys..................... 696
```

FIG. 18 (CONT.[1])

```
932 ATCGGATGAAGAGGACAGTGTGCAGTCGCTCCATGGGGTGAGACATCCTC 981
                  :::   :::      |||||||
697 ........................ProValGlySerGlyAlaHisProG 705

      982 AG......GATCCATGTTCCAGTGATGCTGGGAGA 1010
          ||      ||||||   ::::::|||:::   |||
       705 lnLeuGlyAspProMetGlnAlaAspGlyProArg 716
```

FIG. 18 (CONT.[2])

```
cdinLM_var1     1 TFPLSLIASPFCWTFLRLSISPSFPRVLFPPFSSSHLRPPFLPSFPAHRCFLALLRPRSS
cdinLM_var2     1 TFPLSLIASPFCWTFLRLSISPSFPRVLFPPFSSSHLRPPFLPSFPAHRCFLALLRPRSS
cdinLM_var4     1 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
cdinLM_var5     1 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
cdinLM_var3     1 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
cdinLM_var6     1 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

cdinLM_var1    61 SRPPGVCGLICGPCASVSFSSPFLPTPLPDQRPDPGERMVPEVRVLSSELFLALLWFPLD
cdinLM_var2    61 SRPPGVCGLICGPCASVSFSSPFLPTPLPDQRPDPGERMVPEVRVLSSELFLALLWFPLD
cdinLM_var4     1 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~DKVFGLEFPGTNNA
cdinLM_var5     1 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~DKVFGLEFPGTNNA
cdinLM_var3     1 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
cdinLM_var6     1 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~DKVFGLEFPGTNNA

cdinLM_var1   121 SHARARPDNFCLFHGKKYSEGESWHFYLEFQGLMYCLRCTCSEGNHVSCYRLHFTFVHCT
cdinLM_var2   121 SHARARPDNFCLFHGKKYSPGESWHFYLEFQGLMYCLRCTCSEGNHVSCYKLLHGTLVHCT
cdinLM_var4    15 LVGLFGPDHFCLFHGKKYSPGESWHFYLEFQGLMYCLRCTCSEGNHVSCYRLHFTFVHCT
cdinLM_var5    15 LVGLFGPDHFCLFHGKKYSPGESWHFYLEFQGLMYCLRCTCSEGNHVSCYKLHFTFVHCT
cdinLM_var3     1 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~IS[illegible]W[illegible]QMQNHQKSGLVNFSKD
cdinLM_var6    15 LVGLFGPDHFCLFHGKKYSPGESWHFYLEFQGLMYCLRCTCSEGNHVSCYKLHFTFPVHCT
```

FIG. 19

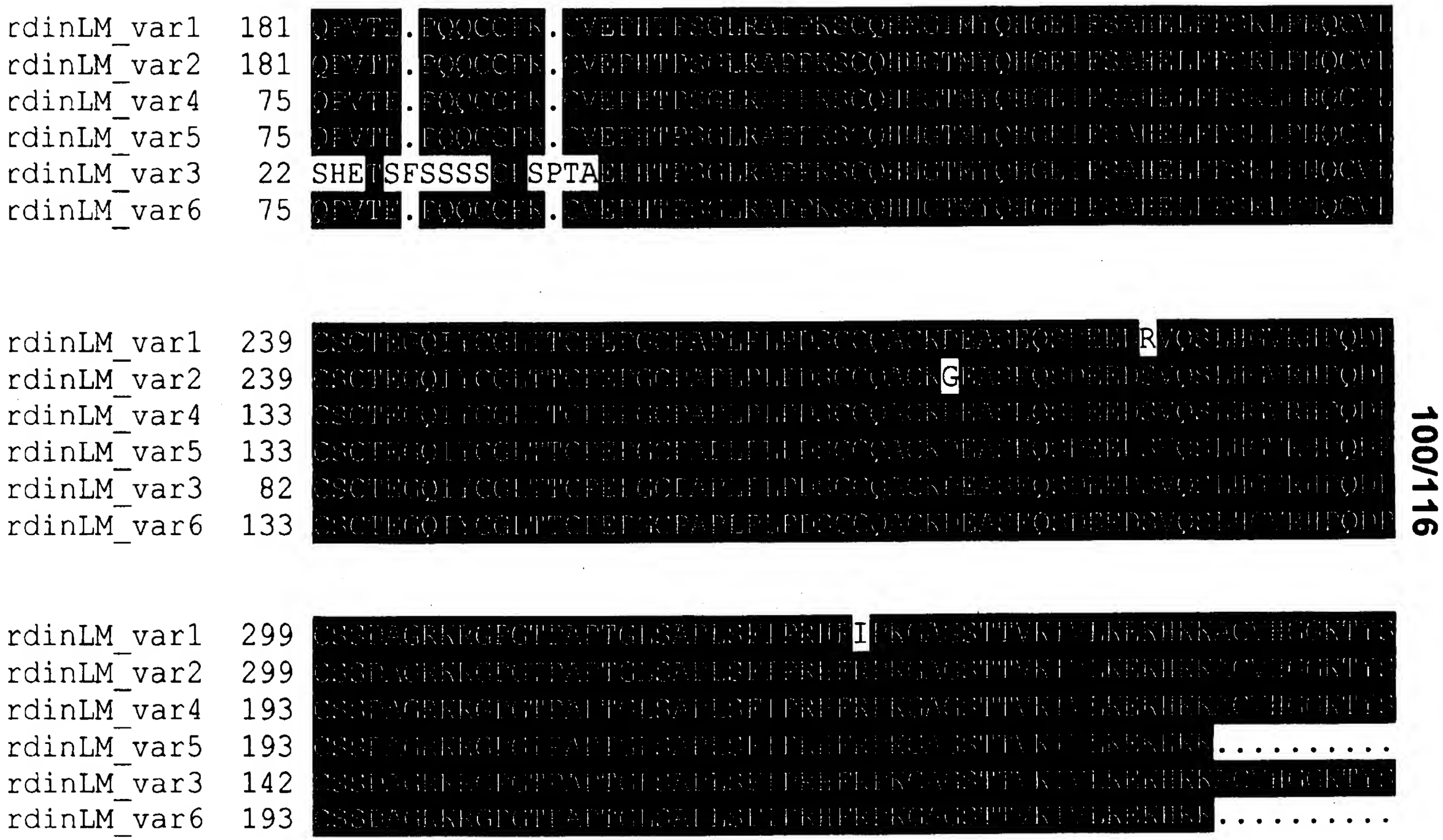

rdinLM_var1 181
rdinLM_var2 181
rdinLM_var4 75
rdinLM_var5 75
rdinLM_var3 22
rdinLM_var6 75
rdinLM_var1 239
rdinLM_var2 239
rdinLM_var4 133
rdinLM_var5 133
rdinLM_var3 82
rdinLM_var6 133
rdinLM_var1 299
rdinLM_var2 299
rdinLM_var4 193
rdinLM_var5 193
rdinLM_var3 142
rdinLM_var6 193

FIG. 19 (CONT.[1])

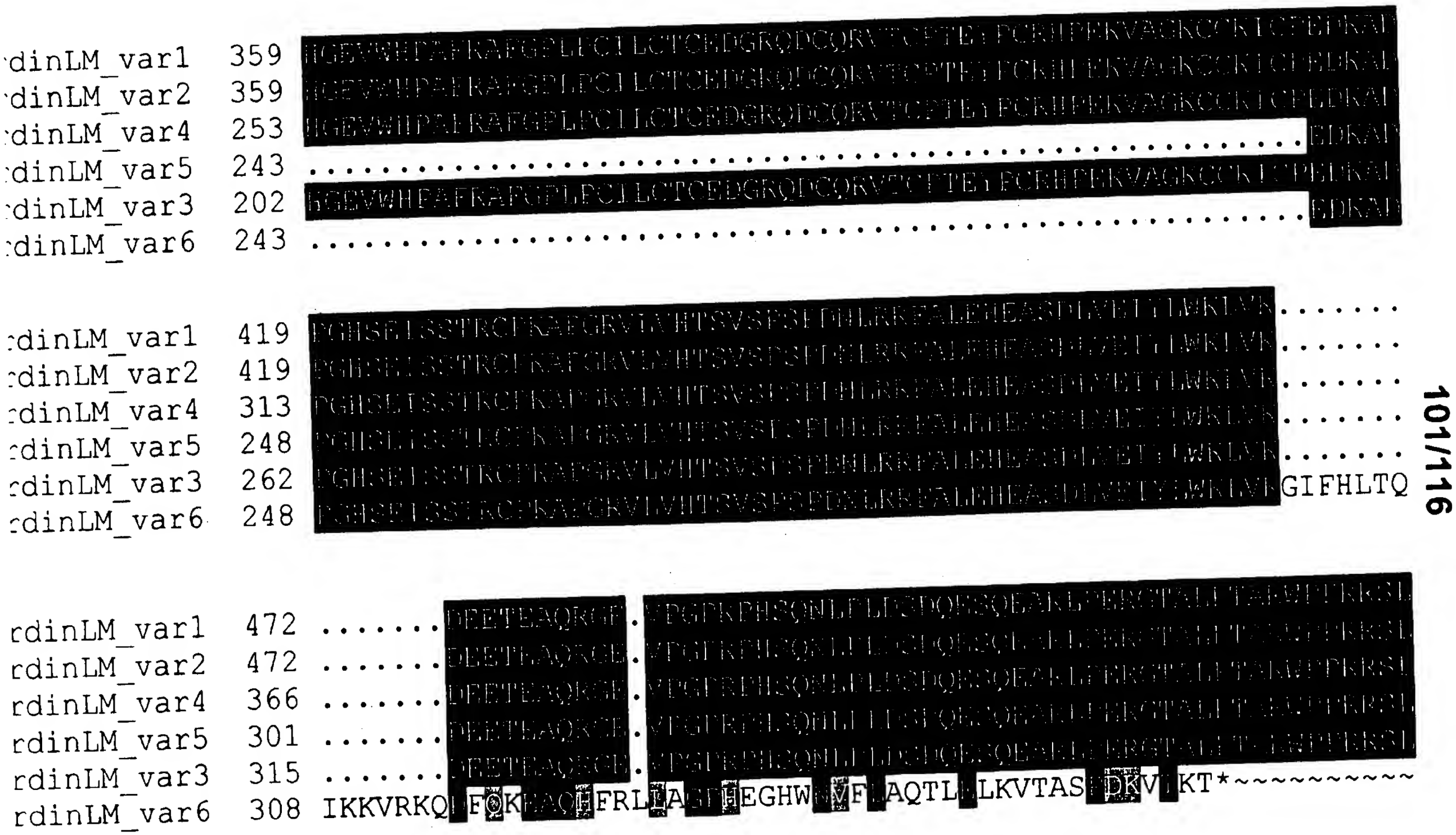

dinLM_var1 359
dinLM_var2 359
dinLM_var4 253
dinLM_var5 243
dinLM_var3 202
dinLM_var6 243
dinLM_var1 419
dinLM_var2 419
dinLM_var4 313
dinLM_var5 248
dinLM_var3 262
dinLM_var6 248
rdinLM_var1 472
rdinLM_var2 472
rdinLM_var4 366
rdinLM_var5 301
rdinLM_var3 315
rdinLM_var6 308
IKKVRKQ F K A H FRL A HEGHW F AQTL LKVTAS DKV KT*~~~~~~~~~~

FIG. 19 (CONT.^{2})

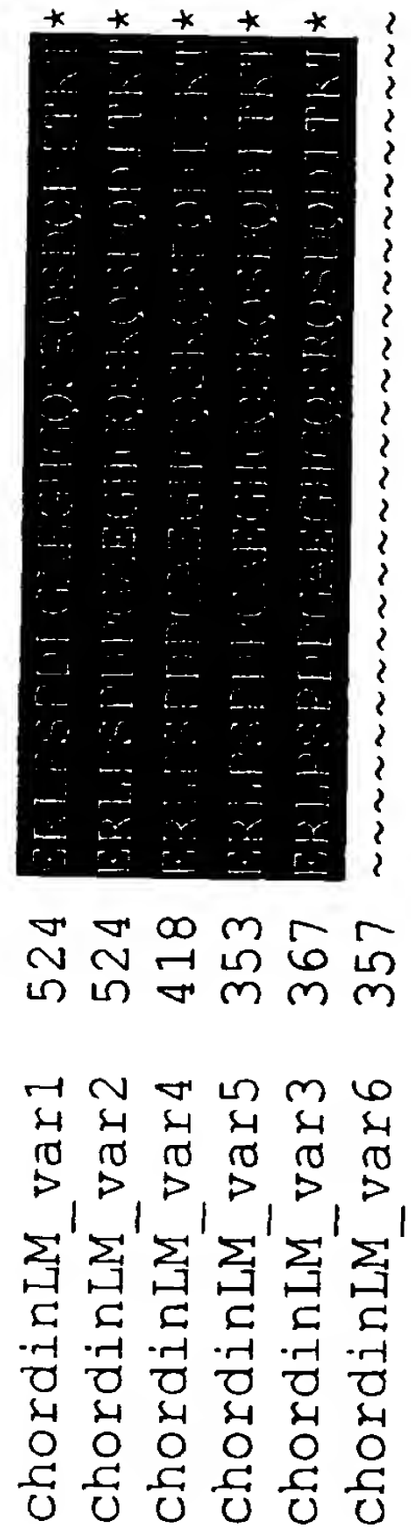

chordinLM_var1
chordinLM_var2
chordinLM_var4
chordinLM_var5
chordinLM_var3
chordinLM_var6
524
524
418
353
367
357

FIG. 19 (CONT.3)

```
425 AGACATTCC...............CAGGATCCATGCTCGGAGAGGAGAGG 459
    |||||||||                |||||||      :::|||:::::
  1 ArgHisSerTyrHisArgSerHisTyrAspProProProSerArgGlnAl 17

460 CCCCAGCACGCCAGCCCCCACCAGCCTCAGCTCCCCTCTGGGCTTCATCC 509
    :   :::                 :::      |||      |||   :::
 17 aGlyGlyLeuSerArgPheProGlyAlaArgSerHisArgGlyAlaLeuM 34

510 NTCGCCACTTCCAGTCAGTAGGAATGGGCAGCACAACCATCAAGATTATC 559
        |||:::   |||          |||   :::::::|||:::
 34 etAspSerGlnGlnAlaSerGly.........ThrIleValGlnIleVal 47

560 TTGAAGGAGAAACATAAA......AAAGCTTGCACACACAATGGGAAGAC 603
    :::::::::|||||||||      :::::::|||:::   ||||||||||
 48 IleAsnAsnLysHisLysHisGlyGlnValCysValSerAsnGlyLysTh 64

604 ATACTCCCATGGGGAGGTGTGGCACCCCACTGTGCTCTCCTTTGGCCCCA 653
    ||||||||||||||||   |||||||||:::::   :::|||||||   :
 64 rTyrSerHisGlyGluSerTrpHisProAsnLeuArgAlaPheGlyIleV 81
```

FIG. 20

```
654 TGCCCTGCATCCTGTGCACATGTATTGATGGCTACCAGGACTGCCACCGT 703
    ::   |||:::|||||||||||||||          |||:::|||   :::
 81 alGluCysValLeuCysThrCysAsnValThrLysGlnGluCysLysLys 97

704 GTGACCTGCCCCACCCAATATCCCTGCAGTCAACCCAAGAAAGTGGCTGG 753
    :::   ||||||:::::|||||||||:::   |||:::|||:::   ||
 98 IleHisCysProAsnArgTyrProCysLysTyrProGlnLysIleAspGl 114

754 GAAGTGCTGCAAGATCTGCCCAGAGGACGAGGCGGAAGATGACCACAGTG 803
       ||||||||||||||::||||||      :::|||:::::::::
114 yLysCysCysLysValCysProGlyLysLysAlaLysGluGlu..... 128

804 AGGTCATTTCCACCCGGTGTCCCAAGGTACCAGGCCAG..... 841
                    :::|||||||||
129 .........................LeuProGlyGlnSerPheAspAsn 136

842 ............................TTCCAGGTGTAC...ACGTTGGC 861
                  :::   ||||||   :::::::
137 LysGlyTyrPheCysGlyGluGluThrMetProValTyrGluSerValPh 153
```

FIG. 20 (CONT.[1])

```
862 ATCTCCAAGCCCAGACAGCCTACACCGCTTTGTCCTGGAGCATGAAGCCT 911
        :::   ::::::   :::::::::::::||||||   |||
153 eMetGluAspGlyGluThrThrArgLysIleAlaLeuGluThrGluArgP 170

912 CTGACCAGGTAGAGATGTACATTTGGAAGCTGGTGAAAGGAATTTACCAC 961
      |||||||||:::::::::|||   :::   |||||||||   :::
170 roProGlnValGluValHisValTrpThrIleArgLysGlyIleLeuGln 186

962 TTGGTTCAGATCAAGAGAGTCAGGAAGCAAGATTTCCAGAAAGAGGTTCA 1011
        :::|||:::::::::   |||:::   |||:::   |||:::
187 HisPheHisIleGluLysIleSerLysArgMetPheGlu...GluLeuPr 202

1012 GAACTTCCGGCTGCTCACCGGCACCCATGAAGGTTACTGGACCGTTTTCC 1061
       :::|||:::|||:::|||   |||       :::   |||   :::|||
 202 oHisPheLysLeuValThrArgThrThrLeuSerGlnTrpLysIlePheT 219

1062 TA.........GCCCAGATTCCAGAGCTGAAAGTTACAGCCAGCCCAGAC 1102
           |||||||||   ::::::                      :::
 219 hrGluGlyGluAlaGlnIleSerGlnMet.............CysSerSer 231
```

FIG. 20 (CONT.^{2})

```
1103 AAAGTGACCAAGACATTATAGCAAGGACCTAAAGAGTTGCAGATACGAGT 1152
       :::|||   :::|||   +++              ||||||:::
 232 ArgValCysArgThr.................GluLeuGluAsp..... 240

      1153 TTTATTGGTTTTGTTATTATATATTAATAAA 1183
             ||||||   :::||||||:::::::::
       241 ....LeuValLysValLeuTyrLeuGluArg 249
```

FIG. 20 (CONT.[3])

```
 18 CCCACACTGCTCTGCCTACCCACACCA...GCCCCAAGGTCTNAGAAAGC 64
    |||   |||   :::   |||   |||   ||||||      :::|||
673 ProMetLeuProAlaGlyProGlyProGluAlaProValProAlaLysHi 689

 65 CCTGGAGGCTGGCTTGCCA...AATCCTTGTCAGTGTNTTTATTGATTAG 111
       :::   |||   |||   :::|||      |||   :::+++
689 sGlySerProGlyArgProArgAspProAsnThrCysPhePhe..... 703

112 TCTGAGAATATCTTAGACCTCACCCACAAGGTTCTGTGTGGAGC..... 155
       +++         :::   :::|||:::|||:::   |||:::
704 .........GluGlyGlnGlnArgProHisGlyAlaArgTrpAlaProAsn 717

156 .........CTGTGCTCTCTGTCTGTCTGT.........CTGTCTGTCTG 187
         ||||||||||||   :::|||                :::||
718 TyrAspProLeuCysSerLeuCysIleCysGlnArgArgThrValIleCy 734

188 TCTGTCTGTCTGTCTGCCTGCCTCTCTCTGTCTGTCTCCGTCTGTCTCTG 237
       |      |||      |||     .   |||      |||
734 sAspProValValCysProProProSerCysProHisPro..... 747
```

FIG. 21

```
238 TCTCTCTGTCTGTCTCTGTCTG.TCTCTTTCTCTCTGTCTCTCTCTGTGT 286
          |||   :::|||:::   |||       :::|||
748 ................ValGlnAlaLeuAspGlnCysCysProValCys 758

287 CTCTGTCTCTGTCTCTGTCTCTCTCTCTCTCAGAAGTCCTCTAGCCTT 336
                           |||
759 ..................................ProGluLysGl 762

337 CTCTAGCAGGCGTCTC..............ATGCAGCCTGGT...TGGT 368
      ||||||   |||                 :::::||||||      |
762 nArgSerArgAspLeuProSerLeuProAsnLeuGluProGlyGluGlyC 779

369 GT............TCCCAGCTGTGGCCTATCCCACAGACAGCTCCACAT 406
    ||            ::::::   |||           |||     |||
779 ysTyrPheAspGlyAspArgSerTrpArgAlaAlaGlyThrArgTrpHis 795

407 CCT.................................GCTTGCTGTTC 420
    |||                 .               :::   |||::
796 ProValValProProPheGlyLeuIleLysCysAlaValCysThrCysLy 812
```

FIG. 21 (CONT.[1])

```
421 GCAGAGACATTCCCAGGATCCATGCTCGGAGAGGAGAGGCCCC..... 463
      :          ::::::     |||::::::   :::   |||
812 sGlyAlaThrGlyGluValHisCysGluLysValGlnCysProArgLeuA 829

464 ....................AGCACGCCAGCCCCCACCAGCCTCAGCTCC 493
                 ::::::|||:::             :::
829 laCysAlaGlnProValArgAlaAsnProThrAspCysCysLysGlnCys 845

494 CCTCTGGGCTTCATCCNTCGCCACTTCCAGTCAGTAGGAATGGGCAGCAC 543
    |||:::|||                                 ::::::||
846 ProValGly.................................SerGlyTh 851

544 AACCATCAAGATT......ATCTTGAAGGAGAAACATAAAAAAGCTTGCA 587
     |:::   |||:::         ::::::              ::::::|||
851 rAsnAlaLysLeuGlyAspProMetGlnAlaAspGlyProArgGlyCysA 868

588 CACACAATGGGAAGACATACTCCCATGGGGAGGTGTGGCACCCCACTGTG 637
       |||:::   :::   :::::::::   |||||||||:::|||
868 rgPheAlaGlyGlnTrpPheProGluAsnGlnSerTrpHisProSerVal 884
```

FIG. 21 (CONT.[2])

```
638 CTCTCCTTTGGCCCCATGCCCTGCATCCTGTGCACATGTATTGATGGCTA 687
        ||||||   |||   ||||||   |||   |||       |||
885 ProProPheGlyGluMetSerCysIleThrCysArgCysGlyAlaGlyVa 901

688 CCAGGACTGCCACCGTGTGACCTGCCCCACCCAATATCCCTGCAGTCAAC 737
        |||:::|||      |||                 |||::::::
901 lProHisCysGluArgAspAspCysSerProProLeuSerCysGlySerG 918

738 CCAAGAAAGTGGCTGGGAAGTGCTGCAAGATCTGC.............. 772
           |||:::     ::::::||||||:::   |||
918 lyLysGlu......SerArgCysCysSerHisCysThrAlaGlnArgSer 932

 773 ..................CCAGAGGACGAGGCGGAAGATGACCACAGT 802
              ||||||   |||   |||   :::||||||
 933 SerGluThrArgThrLeuProGluLeuGluLysGluAlaGluHisSer 948
```

FIG. 21 (CONT.[3])

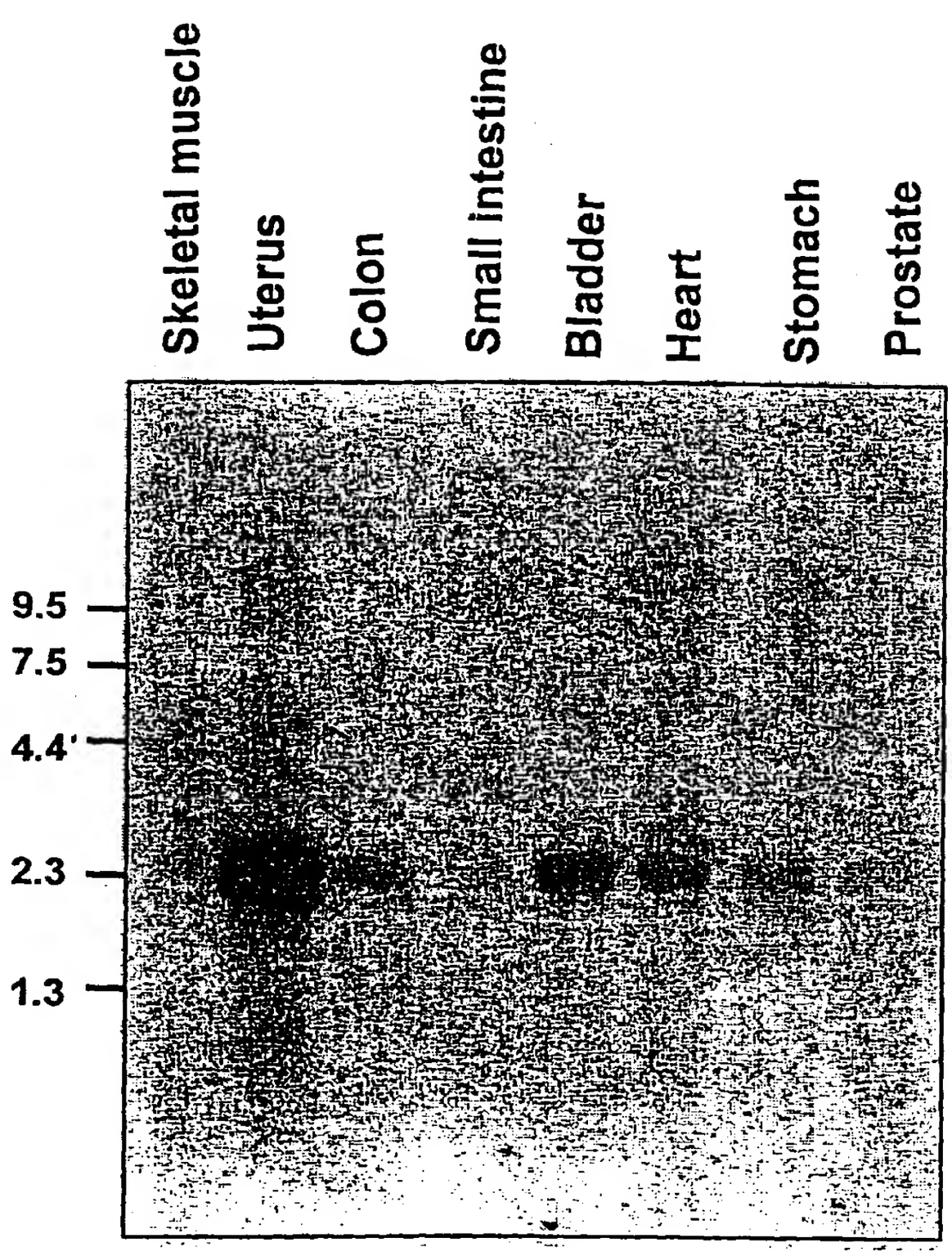

Skeletal muscle
Uterus
Colon
Small intestine
Bladder
Heart
Stomach
Prostate
9.5
7.5
4.4
2.3
1.3

Fig. 22

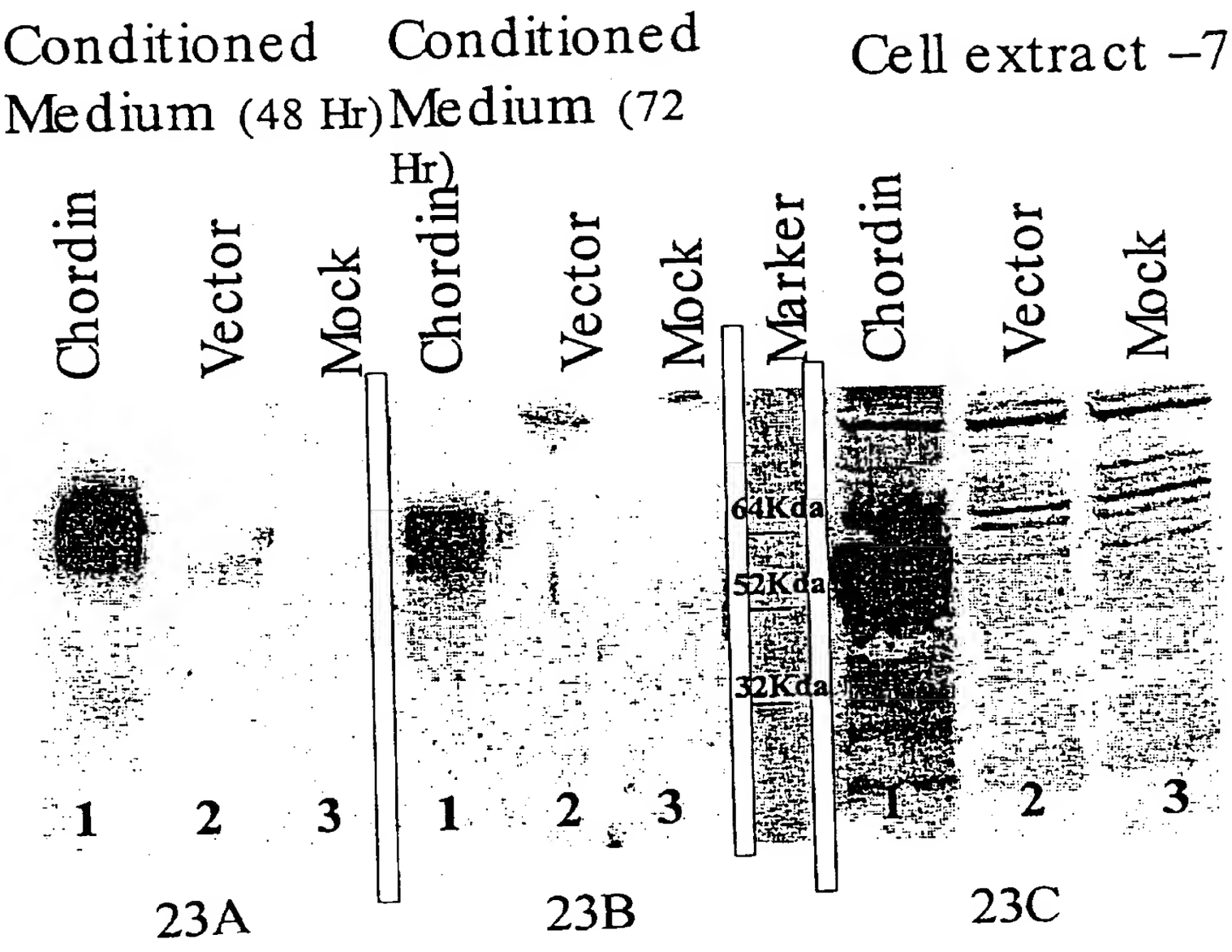

Conditioned Medium (48 Hr)
Conditioned Medium (72 Hr)
Cell extract –72Hr
Chordin
Vector
Mock
Chordin
Vector
Mock
Marker
Chordin
Vector
Mock
64Kda
52Kda
32Kda
1
2
3
1
2
3
1
2
3
23A
23B
23C

Fig. 23

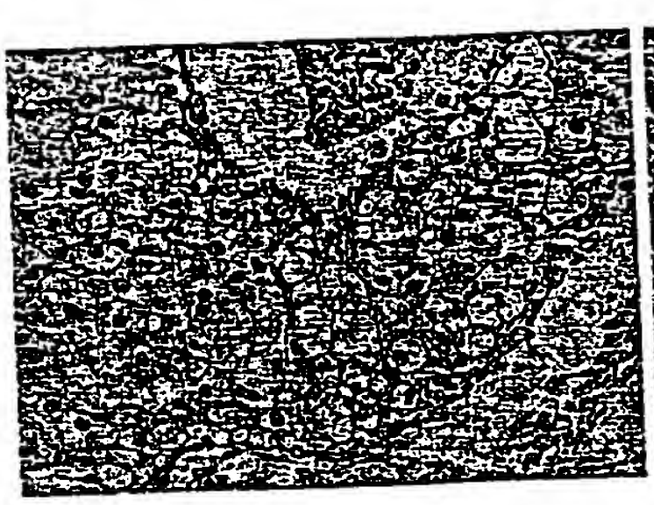


Fig.24A

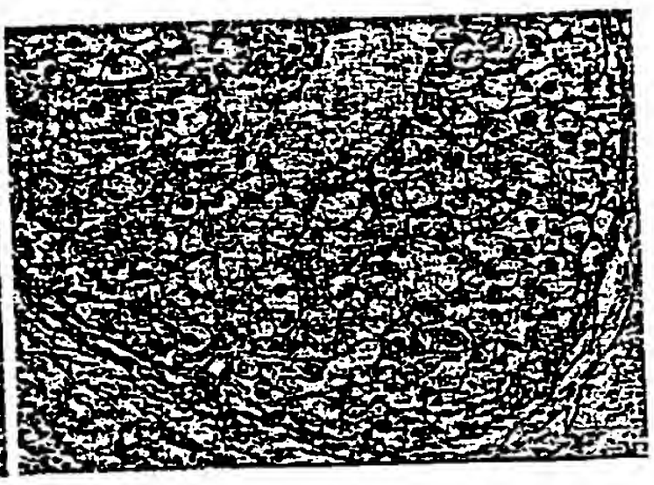


Fig.24A'

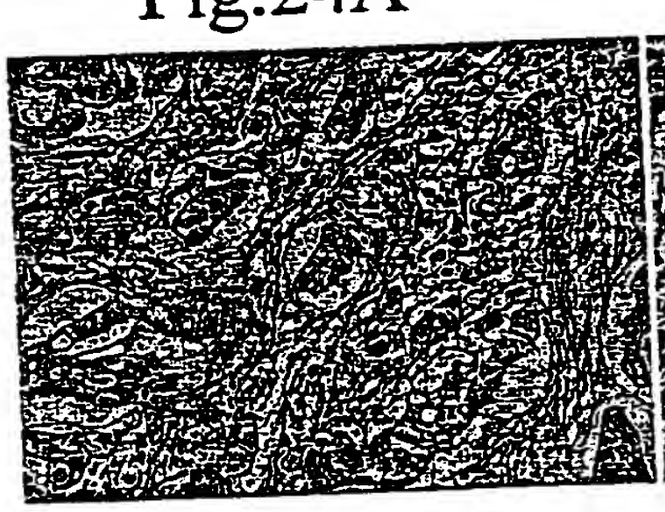


Fig.24B

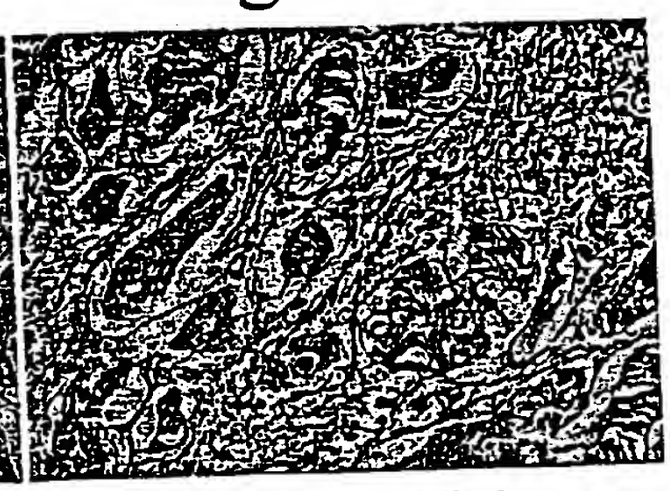


Fig.24B'

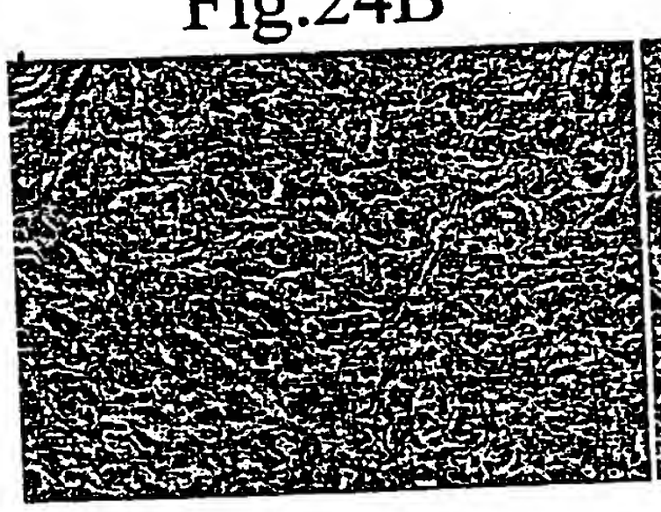


Fig.24C

Fig.24C'

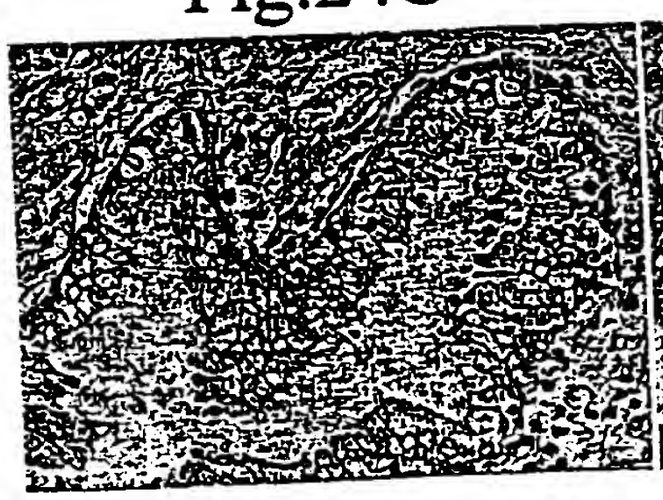


Fig.24D

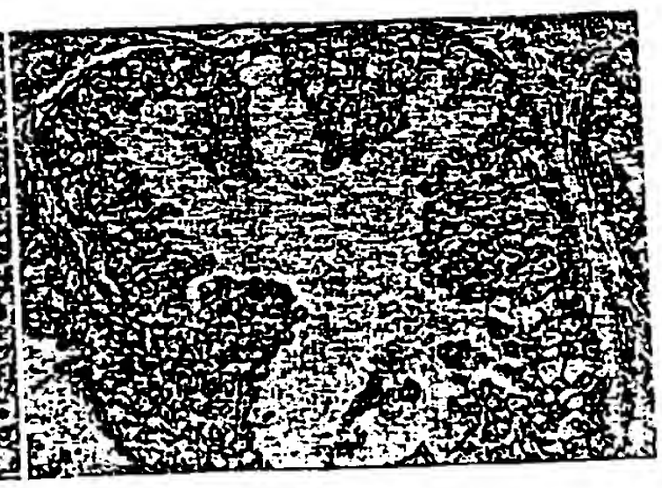


Fig.24D'

Fig.24E

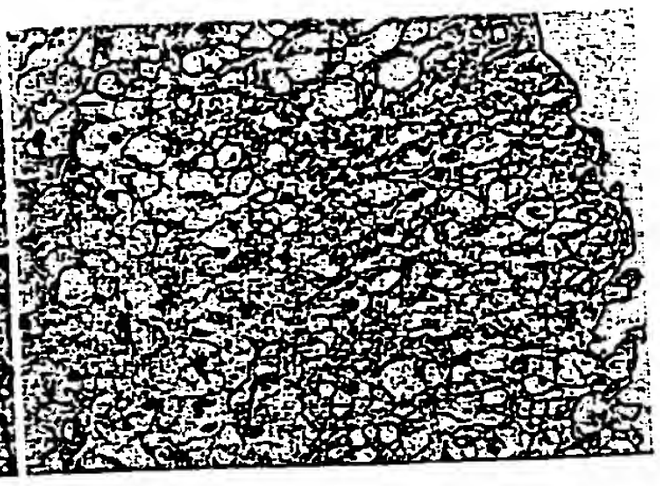


Fig.24E'

Fig.24F

Fig.24F'

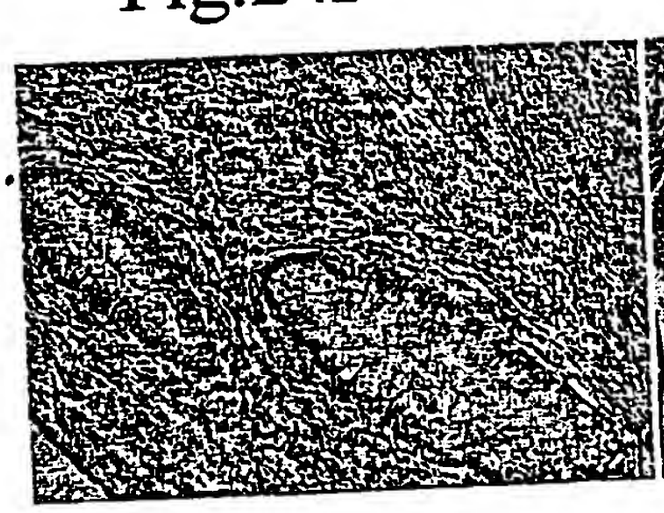


Fig.24G

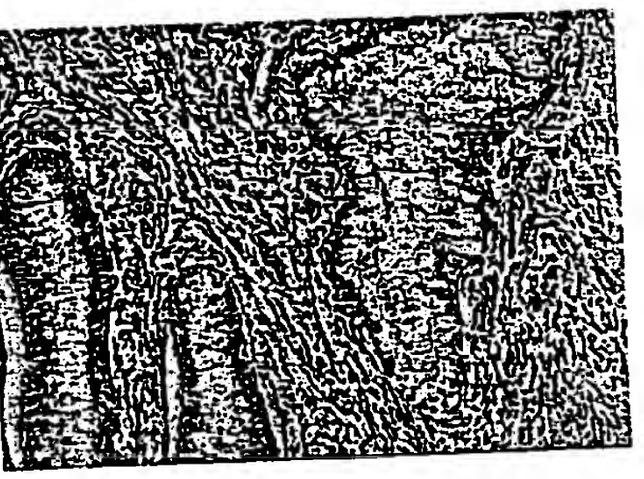


Fig.24G'

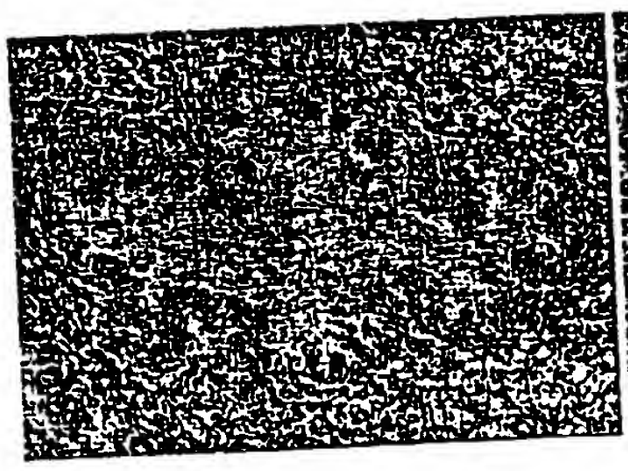


Fig.24H

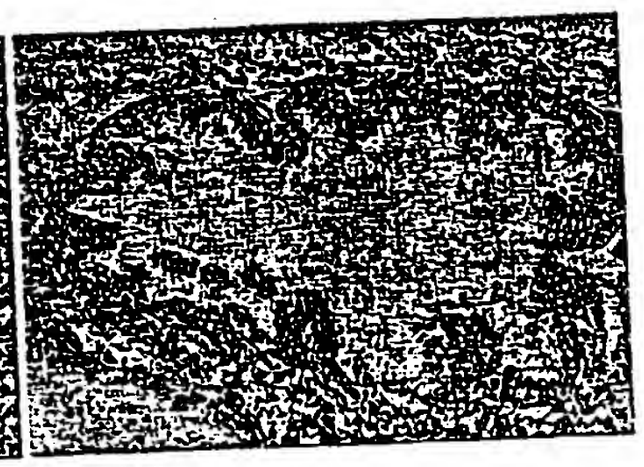


Fig.24H'

Fig.24I Fig.24I'

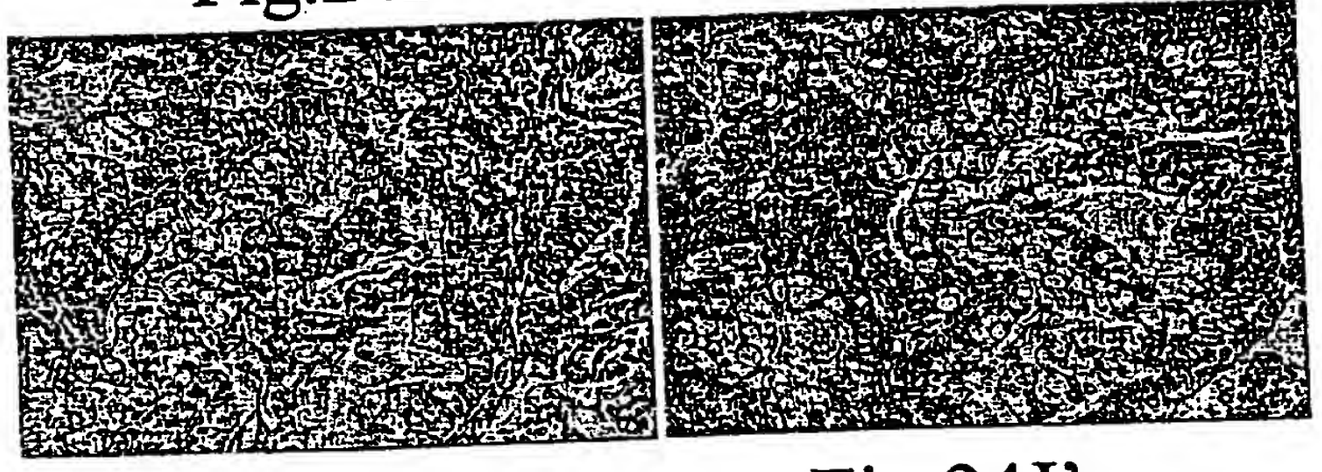


Fig.24J Fig.24J'

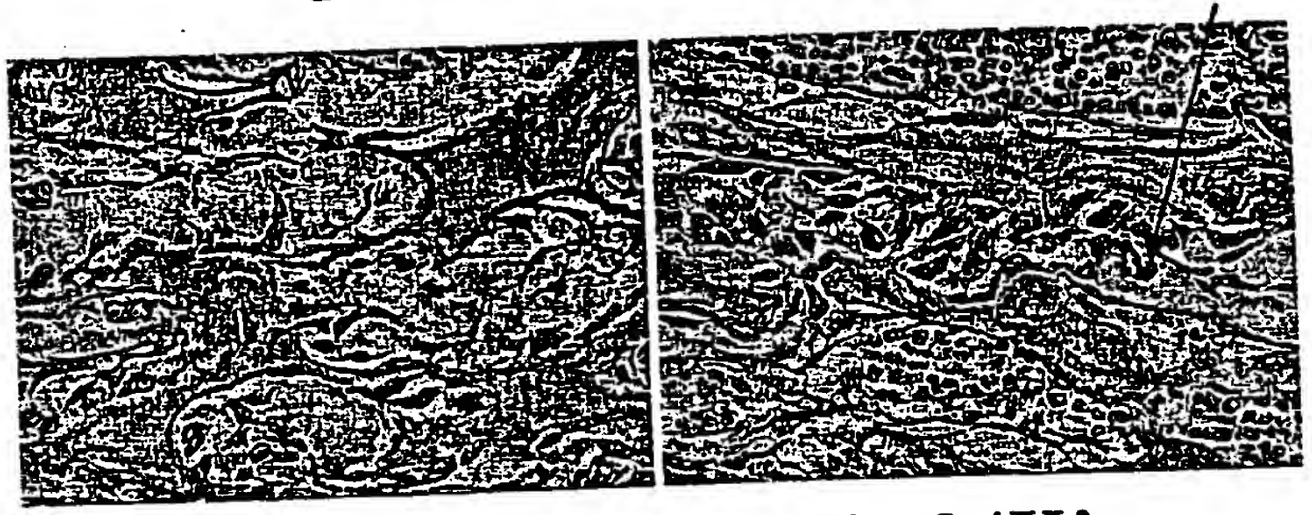


Fig.24K Fig.24K'

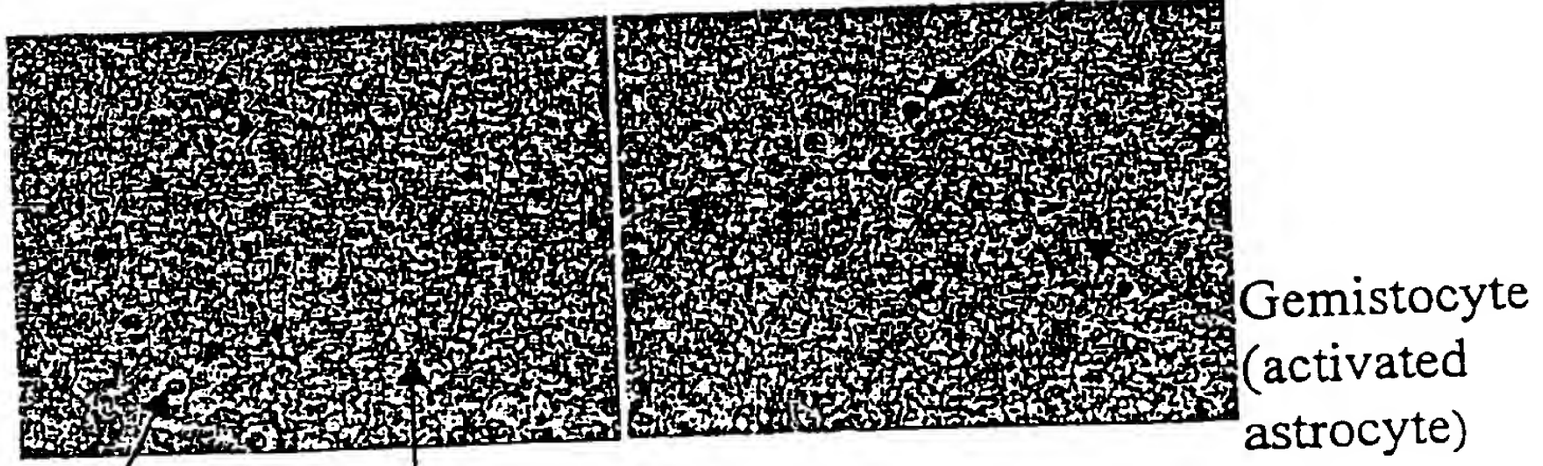
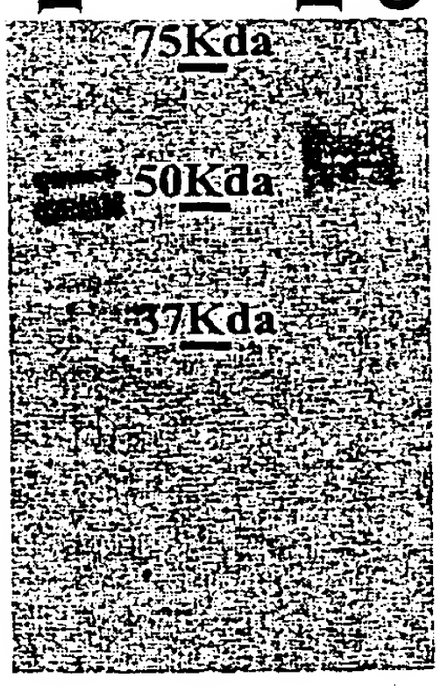


Fig.24L Fig.24L'

25A



25B

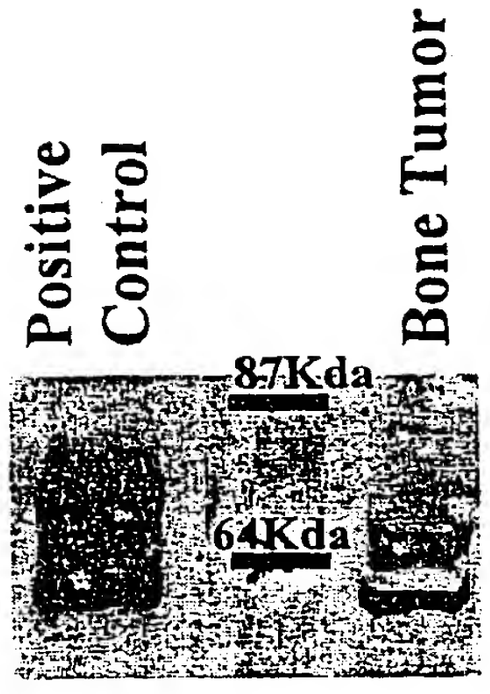


Fig. 25

SEQUENCE LISTING

<110> COMPUGEN LTD

<120> CHORDIN - LIKE HOMOLOGUE

<130> 1292754 - COMPUGEN

<140>
<141>

<150> IL 132846
<151> 1999-11-10

<150> IL 133767
<151> 1999-12-28

<160> 22

<170> PatentIn Ver. 2.1

<210> 1
<211> 1281
<212> DNA
<213> Homo sapiens

<400> 1
cagggagcca caaggcctga tgtactgcct gcgctgtacc tgctcagagg gcgcccatgt 60
gagttgttac cgcctccact gtccgcctgt ccactgcccc cagcctgtga cggagccaca 120
gcaatgctgt cccaagtgtg tggaacctca cactccctct ggactccggg ccccaccaaa 180
gtcctgccag cacaacggga ccatgtacca acacggagag atcttcagtg cccatgagct 240
gttcccctcc cgcctgccca accagtgtgt cctctgcagc tgcacagagg gccagatcta 300
ctgcggcctc acaacctgcc ccgaaccagg ctgcccagca cccctcccgc tgccagactc 360
ctgctgccaa gcctgcaaag atgaggcaag tgagcaatcg gatgaagagg acagtgtgca 420
gtcgctccat ggggtgagac atcctcagga tccatgttcc agtgatgctg ggagaaagag 480
aggcccgggc accccagccc ccactggcct cagcgccccc ctgagcttca tccctcgcca 540
cttcagaccc aagggagcag gcagcacaac tgtcaagatc gtcctgaagg agaaacatan 600
gaaagcctgt gtgcatggcg ggaagacgta ctcccacggg gaggtgtggc acccggcctt 660
ccgtgccttc ggcccttgcc catgcatcct atgcacctgt gaggatggcc gccaggactg 720
ccagcgtgtg acctgtccca cgaagtaccc ctgccgtcac cccgagaaag tggctgggaa 780
gtgctgcaag atttgcccag aggacaaagc agaccctggc cacagtgaga tcagttctac 840
caggtgtccc aaggcaccgg gccgggtcct cgtccacaca tcggtatccc caagcccaga 900
caacctgcgt cgctttgccc tggaacacga ggcctcggac ttggtggaga tctacctctg 960
gaagctggta aaagatgagg aaactgaggc tcagagaggt gaagtacctg gcccaaggcc 1020
acacagccag aatttccact tgactcagat caagaaagtc aggaagcaag acttccagaa 1080
agaggcacag cacttccgac tgctcgctgg cccccacgaa ggtcactgga acgtcttcct 1140
agcccagacc ctggagctga aggtcacggc cagtccagac aaagtgacca agacataaca 1200
aagacctaac agttgcagat atgagctgta taattgttgt tattatatat taataaataa 1260
gaagttgcat aaccatcaaa a 1281

<210> 2
<211> 1722
<212> DNA
<213> Homo sapiens

<400> 2
accttccccc tttctttgat cgcctctccc ttctgctgga ccttccttcg tctctccatc 60
tctccctcct ttccccgcgt tctctttcca cctttctctt cttcccacct tagacctccc 120
ttcctgccct cctttcctgc ccaccgctgc ttcctggccc ttctccgacc ccgctctagc 180
agcagacctc ctggggtcat gtgggttgat ctgtggcccc tgtgnctccg tgtccttttc 240

gtctcccgtc ctcccgactc cgctcccgga ccagcggcct gaccctgggg aaaggatggt 300
tcccgaggtg agggtcctct cctccttgct gggactcgcg ctgctctggt tccccctgga 360
ctcccacgct cgagcccgcc cagacatgtt ctgccttttc catgggaaga gatactcccc 420
cggcgagagc tggcaccct acttggagcc acaaggcctg atgtactgcc tgcgctgtac 480
ctgctcagag ggcgcccatg tgagttgtta ccgcctccac tgtccgcctg tccactgccc 540
ccagcctgtg acggagccac agcaatgctg tcccaagtgt gtggaacctc acactccctc 600
tggactccgg gccccaccaa agtcctgcca gcacaacggg accatgtacc aacacggaga 660
gatcttcagt gcccatgagc tgttcccctc ccgcctgccc aaccagtgtg tcctctgcag 720
ctgcacagag ggccagatct actgcggcct cacaacctgc cccgaaccag gctgcccagc 780
acccctcccg ctgccagact cctgctgcca agcctgcaaa gatgaggcaa gtgagcaatc 840
ggatgaagag gacagtgtgc agtcgctcca tggggtgaga catcctcagg atccatgttc 900
cagtgatgct gggagaaaga gaggcccggg caccccagcc cccactggcc tcagcgcccc 960
tctgagcttc atccctcgcc acttcagacc caagggagca ggcagcacaa ctgtcaagat 1020
cgtcctgaag gagaaacata ngaaagcctg tgtgcatggc gggaagacgt actcccacgg 1080
ggaggtgtgg cacccggcct tccgtgcctt cggcccttgc ccatgcatcc tatgcacctg 1140
tgaggatggc cgccaggact gccagcgtgt gacctgtccc acgaagtacc cctgccgtca 1200
ccccgagaaa gtggctggga agtgctgcaa gatttgccca gaggacaaag cagaccctgg 1260
ccacagtgag atcagttcta ccaggtgtcc caaggcaccg ggccgggtcc tcgtccacac 1320
atcggtatcc ccaagcccag acaacctgcg tcgctttgcc ctggaacacg aggcctcgga 1380
cttggtggag atctacctct ggaagctggt aaaagatgag gaaactgagg ctcagagagg 1440
tgaagtacct ggcccaaggc cacacagcca gaatttccac ttgactcaga tcaagaaagt 1500
caggaagcaa gacttccaga aagaggcaca gcacttccga ctgctcgctg gcccccacga 1560
aggtcactgg aacgtcttcc tagcccagac cctggagctg aaggtcacgg ccagtccaga 1620
caaagtgacc aagacataac aaagacctaa cagttgcaga tatgagctgt ataattgttg 1680
ttattatata ttaataaata agaagttgca taaccatcaa aa 1722

<210> 3
<211> 1515
<212> DNA
<213> Homo sapiens

<400> 3
agaacagtgc ctggnactga gaagtgctnc atgaaccctc attaaaatgc tggatgaact 60
aactcgggcc ccactgagcg gaagctgagg ctgccggctt taccttcttg ctctggttgg 120
ggctggccca gacactgagg gggctggagg gctgtggtag aggtcatggg agggagggac 180
tcagtcagat gtaggtatca gagggacctc ttagttagct gataagggga atggctggca 240
aggcccaggc cagagcttgg tttaaatatc aagctggggt caaatgcaaa atcatcagaa 300
aagtggcctt gttaatttca gcaaagattc acatgaaacc tcattttctt cttcctcctg 360
cccctccccc actgtggaac ctcacactcc ctctggactc cgggccccac caaagtcctg 420
ccagcacaac gggaccatgt accaacacgg agagatcttc agtgcccatg agctgttccc 480
ctcccgcctg cccaaccagt gtgtcctctg cagctgcaca gagggccaga tctactgcgg 540
cctcacaacc tgccccgaac caggctgccc agcacccctc ccgctgccag actcctgctg 600
ccaagcctgc aaagatgagg caagtgagca atcggatgaa gaggacagtg tgcagtcgct 660
ccatggggtg agacatcctc aggatccatg ttccagtgat gctgggagaa agagaggccc 720
gggcacccca gccccctactg gcctcagcgc ccctctgagc ttcatccctc gccacttcag 780
acccaaggga gcaggcagca caactgtcaa gatcgtcctg aaggagaaac atangaaagc 840
ctgtgtgcat ggcgggaaga cgtactccca cggggaggtg tggcacccgg ccttccgtgc 900
cttcggccct tgcccatgca tcctatgcac ctgtgaggat ggccgccagg actgccagcg 960
tgtgacctgt cccacgaagt acccctgccg tcaccccgag aaagtggctg ggaagtgctg 1020
caagatttgc ccagaggaca aagcagaccc tggccacagt gagatcagtt ctaccaggtg 1080
tcccaaggca ccgggccggg tcctcgtcca cacatcggta tccccaagcc cagacaacct 1140
gcgtcgcttt gccctggaac acgaggcctc ggacttggtg gagatctacc tctggaagct 1200
ggtaaaagat gaggaaactg aggctcagag aggtgaagta cctggcccaa ggccacacag 1260
ccagaatttc cacttgactc agatcaagaa agtcaggaag caagacttcc agaaagaggc 1320
acagcacttc cgactgctcg ctggccccca cgaaggtcac tggaacgtct tcctagccca 1380
gaccctggag ctgaaggtca cggccagtcc agacaaagtg accaagacat aacaaagacc 1440
taacagttgc agatatgagc tgtataattg ttgttattat atattaataa ataagaagtt 1500
gcataaccat caaaa 1515

<210> 4
<211> 1890
<212> DNA
<213> Homo sapiens

<400> 4
tagnancgnc cgccaggggn tgnattcgcc cttactcann atagggntna acggccgccc 60
ggncaggtgc caaatggaca aataaaagga aacaagcatg attgtnaggn cagaggaggg 120
tggnattnag tcagaagact ggtgctgtca tcgctgcntg gtgactgact tgctgtgtgg 180
ccntcaggtg taacttaccc tctctgggcc tcatttgtct aatcataata attaacgttg 240
ataccatgat ataaatctgt acagcatttc actgcttgat tccctaactg ccctgtgaga 300
taagcgttaa ggctcagaga cagtggcatg cccagtgttg cacagtaagt gtgtggtaaa 360
gccgagattc aaactcagac cttctggccc cttgcctagg agagcatgcc cagttgtcta 420
gcagattctc ttttgtctga gtggcccaga tgacatctct tttagagcta gaaagaagga 480
gaaatgagac agggtctttg ggctggagcc tcctgggact aacatggcac tggtcggttt 540
gccaggccca gacatgttct gccttttcca tgggaagaga tactcccccg gcgagagctg 600
gcacccctac ttggagccac aaggcctgat gtactgcctg cgctgtacct gctcagaggg 660
cgcccatgtg agttgttacc gcctccactg tccgcctgtc cactgccccc agcctgtgac 720
ggagccacag caatgctgtc ccaagtgtgt ggaacctcac actccctctg gactccgggc 780
cccaccaaag tcctgccagc acaacgggac catgtaccaa cacggagaga tcttcagtgc 840
ccatgagctg ttcccctccc gcctgcccaa ccagtgtgtc ctctgcagct gcacagaggg 900
ccagatctac tgcggcctca caacctgccc cgaaccaggc tgcccagcac ccctcccgct 960
gccagactcc tgctgccaag cctgcaaaga tgaggcaagt gagcaatcgg atgaagagga 1020
cagtgtgcag tcgctccatg gggtgagaca tcctcaggat ccatgttcca gtgatgctgg 1080
gagaaagaga ggcccgggca ccccagcccc cactggcctc agcgcccctc tgagcttcat 1140
ccctcgccac ttcagaccca agggagcagg cagcacaact gtcaagatcg tcctgaagga 1200
gaaacatang aaagcctgtg tgcatggcgg gaagacgtac tcccacgggg aggtgtggca 1260
cccggccttc cgtgccttcg gcccttgccc atgcatccta tgcacctgtg aggatggccg 1320
ccaggactgc cagcgtgtga cctgtcccac gaagtacccc tgccgtcacc ccgagaaagt 1380
ggctgggaag tgctgcaaga tttgcccaga ggacaaagca gaccctggcc acagtgagat 1440
cagttctacc aggtgtccca aggcaccggg ccgggtcctc gtccacacat cggtatcccc 1500
aagcccagac aacctgcgtc gctttgccct ggaacacgag gcctcggact tggtggagat 1560
ctacctctgg aagctggtaa aagatgagga aactgaggct cagagaggtg aagtacctgg 1620
cccaaggcca cacagccaga atttccactt gactcagatc aagaaagtca ggaagcaaga 1680
cttccagaaa gaggcacagc acttccgact gctcgctggc ccccacgaag gtcactggaa 1740
cgtcttccta gcccagaccc tggagctgaa ggtcacggcc agtccagaca aagtgaccaa 1800
gacataacaa agacctaaca gttgcagata tgagctgtat aattgttgtt attatatatt 1860
aataaataag aagttgcata accatcaaaa 1890

<210> 5
<211> 1722
<212> DNA
<213> Homo sapiens

<400> 5
accttccccc tttctttgat cgcctctccc ttctgctgga ccttccttcg tctctccatc 60
tctccctcct ttccccgcgt tctctttcca cctttctctt cttcccacct tagacctccc 120
ttcctgccct cctttcctgc ccaccgctgc ttcctggccc ttctccgacc ccgctctagc 180
agcagacctc ctggggtctg tgggttgatc tgtggccccc gtgcctccgt gtccttttcg 240
tctcccttcc tcccgactcc gctcccggac cagcggcctg accctgggga aaggatggtt 300
cccgaggtga gggtcctctc ctccttgctg ggactcgcgc tgctctggtt ccccctggac 360
tcccacgctc gagcccgccc agacatgttc tgccttttcc atgggaagag atactccccc 420
ggcgagagct ggcacccccta cttggagcca caaggcctga tgtactgcct gcgctgtacc 480
tgctcagagg gcgcccatgt gagttgttac cgcctccact gtccgcctgt ccactgcccc 540
cagcctgtga cggagccaca gcaatgctgt cccaagtgtg tggaacctca cactccctct 600
ggactccggg ccccaccaaa gtcctgccag cacaacggga ccatgtacca acacggagag 660
atcttcagtg cccatgagct gttcccctcc cgcctgccca accagtgtgt cctctgcagc 720
tgcacagagg gccagatcta ctgcgggctc acaacctgcc ccgaaccagg ctgcccagca 780
ccccctcccgc tgccagactc ctgctgccag gcctgcaaag atgaggcaag tgagcaatcg 840
gatgaagagg accgtgtgca gtcgctccat ggggtgagac atcctcagga tccatgttcc 900

agtgatgctg ggagaaagag aggcccgggc accccagccc ccactggcct cagcgccccc 960
ctgagcttca tccctcgcca cttcataccc aagggagcag gcagcacaac tgtcaagatc 1020
gtcctgaagg agaaacataa gaaagcctgt gtgcatggcg ggaagacgta ctcccacggg 1080
gaggtgtggc acccggcctt ccgtgccttc ggcccccttgc cctgcatcct atgcacctgt 1140
gaggatggcc gccaggactg ccagcgtgtg acctgtccca ccgagtaccc ctgccgtcac 1200
cccgagaaag tggctgggaa gtgctgcaag atttgcccag aggacaaagc agaccctggc 1260
cacagtgaga tcagttctac caggtgtccc aaggcaccgg gccgggtcct cgtccacaca 1320
tcggtatccc caagcccaga caacctgcgt cgctttgccc tggaacacga ggcctcggac 1380
ctggtggaga tctacctctg gaagctggta aaagatgagg aaactgaggc tcagagaggt 1440
gaagtacctg gcccaaggcc acacagccag aatcttccac ttgactcaga tcaagaaagt 1500
caggaagcaa gacttccaga aagaggcaca gcacttccga ctgctcgctg gcccccacga 1560
aggtcactgg aacgtcttcc tagcccagac cctggagctg aaggtcacgg ccagtccaga 1620
caaagtgacc aagacataac aaagacctaa cagttgcaga tatgagctgt ataattgctg 1680
ttattatata ttaataaata agaagttgca taaccatcaa aa 1722

<210> 6
<211> 1722
<212> DNA
<213> Homo sapiens

<400> 6
accttccccc tttctttgat cgcctctccc ttctgctgga ccttccttcg tctctccatc 60
tctccctcct ttccccgcgt tctctttcca cctttctctt cttcccacct tagacctccc 120
ttcctgccct cctttcctgc ccaccgctgc ttcctggccc ttctccgacc ccgctctagc 180
agcagacctc ctggggtctg tgggttgatc tgtggccccct gtgcctccgt gtccttttcg 240
tctcccttcc tcccgactcc gctcccggac cagcggcctg accctgggga aaggatggtt 300
cccgaggtga gggtcctctc ctccttgctg ggactcgcgc tgctctggtt ccccctggac 360
tcccacgctc gagcccgccc agacatgttc tgccttttcc atgggaagag atactccccc 420
ggcgagagct ggcaccccta cttggagcca caaggcctga tgtactgcct gcgctgtacc 480
tgctcagagg gcgcccatgt gagttgttac cgcctccact gtccgcctgt ccactgcccc 540
cagcctgtga cggagccaca gcaatgctgt cccaagtgtg tggaacctca cactccctct 600
ggactccggg ccccaccaaa gtcctgccag cacaacggga ccatgtacca acacggagag 660
atcttcagtg cccatgagct gttcccctcc cgcctgccca accagtgtgt cctctgcagc 720
tgcacagagg gccagatcta ctgcggcctc acaacctgcc ccgaaccagg ctgcccagca 780
cccctcccgc tgccagactc ctgctgccag gcctgcaaag gtgaggcaag tgagcaatcg 840
gatgaagagg acagtgtgca gtcgctccat ggggtgagac atcctcagga tccatgttcc 900
agtgatgctg ggagaaagag aggcccgggc accccagccc ccactggcct cagcgccccc 960
ctgagcttca tccctcgcca cttcagaccc aagggagcag gcagcacaac tgtcaagatc 1020
gtcctgaagg agaaacataa gaaagcctgt gtgcatggcg ggaagacgta ctcccacggg 1080
gaggtgtggc acccggcctt ccgtgccttc ggcccccttgc cctgcatcct atgcacctgt 1140
gaggatggcc gccaggactg ccagcgtgtg acctgtccca ccgagtaccc ctgccgtcac 1200
cccgagaaag tggctgggaa gtgctgcaag atttgcccag aggacaaagc agaccctggc 1260
cacagtgaga tcagttctac caggtgtccc aaggcaccgg gccgggtcct cgtccacaca 1320
tcggtatccc caagcccaga caacctgcgt cgctttgccc tggaacacga ggcctcggac 1380
ttggtggaga tctacctctg gaagctggta aaagatgagg aaactgaggc tcagagaggt 1440
gaagtacctg gcccaaggcc acacagccag aatcttccac ttgactcaga tcaagaaagt 1500
caggaagcaa gacttccaga aagaggcaca gcacttccga ctgctcgctg gcccccacga 1560
aggtcactgg aacgtcttcc tagcccagac cctggagctg aaggtcacgg ccagtccaga 1620
caaagtgacc aagacataac aaagacctaa cagttgcaga tatgagctgt ataattgttg 1680
ttattatata ttaataaata agaagttgca taaccatcaa aa 1722

<210> 7
<211> 1515
<212> DNA
<213> Homo sapiens

<400> 7
agaacagtgc ctggcactga gaagtgctcc atgaaccctc attaaatgct ggatgaacta 60
actcgggccc cactgagcgg aagctgaggc tgccggcttt accttcttgc tctggttggg 120

gctggcccag acactgaggg ggctggaggg ctgtggtaga ggtcatggga gggagggact 180
cagtcagatg taggtatcag agggacctct tagttagctg ataaggggaa tggctggcaa 240
ggcccaggcc agagcttggt ttaaatatca agctggggtc aaatgcaaaa tcatcagaaa 300
agtggccttg ttaatttcag caaagattca catgaaacct cattttcttc ttcctcctgc 360
ccctccccca ctgcagaacc tcacactccc tctggactcc gggccccacc aaagtcctgc 420
cagcacaacg ggaccatgta ccaacacgga gagatcttca gtgcccatga gctgttcccc 480
tcccgcctgc ccaaccagtg tgtcctctgc agctgcacag agggccagat ctactgcggc 540
ctcacaacct gccccgaacc aggctgccca gcacccctcc cgctgccaga ctcctgctgc 600
caggcctgca aagatgaggc aagtgagcaa tcggatgaag aggacagtgt gcagtcgctc 660
catggggtga gacatcctca ggatccatgt tccagtgatg ctgggagaaa gagaggcccg 720
ggcaccccag cccccactgg cctcagcgcc cctctgagct tcatccctcg ccacttcaga 780
cccaagggag caggcagcac aactgtcaag atcgtcctga aggagaaaca taagaaagcc 840
tgtgtgcatg gcgggaagac gtactcccac ggggaggtgt ggcacccggc cttccgtgcc 900
ttcggccccт tgccctgcat cctatgcacc tgtgaggatg gccgccagga ctgccagcgt 960
gtgacctgtc ccaccgagta cccctgccgt caccccgaga aagtggctgg gaagtgctgc 1020
aagatttgcc cagaggacaa agcagaccct ggccacagtg agatcagttc taccaggtgt 1080
cccaaggcac cgggccgggt cctcgtccac acatcggtat ccccaagccc agacaacctg 1140
cgtcgctttg ccctggaaca cgaggcctcg gacttggtgg agatctacct ctggaagctg 1200
gtaaaagatg aggaaactga ggctcagaga ggtgaagtac ctggcccaag gccacacagc 1260
cagaatcttc cacttgactc agatcaagaa agtcaggaag caagacttcc agaaagaggc 1320
acagcacttc cgactgctcg ctggccccca cgaaggtcac tggaacgtct tcctagccca 1380
gaccctggag ctgaaggtca cggccagtcc agacaaagtg accaagacat aacaaagacc 1440
taacagttgc agatatgagc tgtataattg ttgttattat atattaataa ataagaagtt 1500
gcataaccat caaaa 1515

<210> 8
<211> 1817
<212> DNA
<213> Homo sapiens

<400> 8
ggacaaataa aaaggaaaca agcatgattg tgagggcaga ggagcgtggg actgagtcag 60
gagactggtg ctgtcatcgc tgcctggtga ctgacttgct gtgtggccct caggtgtaac 120
ttaccctctc tgggcctcat ttgtctaatc ataataatta acgctgatac catgatataa 180
atctgtacag catttcactg cttgattccc taactgccct gtgagataag cgttaaggct 240
cagagacagt ggcatgccca gtgttgcaca gtaagtgtgt ggtaaagccg agattcaaac 300
tcagaccttc tggccccttg cctaggagag catgcccagt tgtctagcag attctctttt 360
gcctgagtgg cccagatgac atctctttta gagctagaaa gaaggagaaa tgagacaggg 420
tctttgggct ggagcctcct gggactaaca tggcactggt cggtttgcca ggcccagaca 480
tgttctgcct tttccatggg aagagatact ccccggcga gagctggcac ccctacttgg 540
agccacaagg cctgatgtac tgcctgcgct gtacctgctc agagggcgcc catgtgagtt 600
gttaccgcct ccactgtccg cctgtccact gcccccagcc tgtgacggag ccacagcaat 660
gctgtcccaa gtgtgtggaa cctcacactc cctctggact ccgggcccca ccaaagtcct 720
gccagcacaa cgggaccatg taccaacacg gagagatctt cagtgcccat gagctgttcc 780
cctcccgcct gcccaaccag tgtgtcctct gcagctgcac agagggccag atctactgcg 840
gcctcacaac ctgccccgaa ccaggctgcc cagcacccct cccgctgcca gactcctgct 900
gccaagcctg caaagatgag gcaagtgagc aatcggatga agaggacagt gtgcagtcgc 960
tccatggggt gagacatcct caggatccat gttccagtga tgctgggaga aagagaggcc 1020
cgggcacccc agcccccact ggcctcagcg cccctctgag cttcatccct cgccacttca 1080
gacccaaggg agcaggcagc acaactgtca agatcgtcct gaaggagaaa cataagaaag 1140
cctgtgtgca tggcgggaag acgtactccc acggggaggt gtggcacccg gccttccgtg 1200
ccttcggccc cttgccctgc atcctatgca cctgtgagga tggccgccag gactgccagc 1260
gtgtgacctg tcccaccgag taccctgcc gtcaccccga gaaagtggct gggaagtgct 1320
gcaagatttg cccagaggac aaagcagacc ctggccacag tgagatcagt tctaccaggt 1380
gtcccaaggc accgggccgg gtcctcgtcc acacatcggt atccccaagc ccagacaacc 1440
tgcgtcgctt tgccctggaa cacgaggcct cggacttggt ggagatctac ctctggaagc 1500
tggtaaaaga tgaggaaact gaggctcaga gaggtgaagt acctggccca aggccacaca 1560
gccagaatct tccacttgac tcagatcaag aaagtcagga agcaagactt ccagaaagag 1620
gcacagcact tccgactgct cgctggcccc cacgaaggtc actggaacgt cttcctagcc 1680
cagaccctgg agctgaaggt cacggccagt ccagacaaag tgaccaagac ataacaaaga 1740

cctaacagtt gcagatatga gctgtataat tgttgttatt atatattaat aaataagaag 1800
ttgcataacc atcaaaa 1817

<210> 9
<211> 1622
<212> DNA
<213> Homo sapiens

<400> 9
ggacaaataa aaaggaaaca agcatgattg tgagggcaga ggagcgtggg actgagtcag 60
gagactggtg ctgtcatcgc tgcctggtga ctgacttgct gtgtggccct caggtgtaac 120
ttaccctctc tgggcctcat ttgtctaatc ataataatta acgctgatac catgatataa 180
atctgtacag catttcactg cttgattccc taactgccct gtgagataag cgttaaggct 240
cagagacagt ggcatgccca gtgttgcaca gtaagtgtgt ggtaaagccg agattcaaac 300
tcagaccttc tggccccttg cctaggagag catgcccagt tgtctagcag attctctttt 360
gcctgagtgg cccagatgac atctctttta gagctagaaa gaaggagaaa tgagacaggg 420
tctttgggct ggagcctcct gggactaaca tggcactggt cggtttgcca ggcccagaca 480
tgttctgcct tttccatggg aagagatact cccccggcga gagctggcac ccctacttgg 540
agccacaagg cctgatgtac tgcctgcgct gtacctgctc agagggcgcc catgtgagtt 600
gttaccgcct ccactgtccg cctgtccact gcccccagcc tgtgacggag ccacagcaat 660
gctgtcccaa gtgtgtggaa cctcacactc cctctggact ccgggcccca ccaaagtcct 720
gccagcacaa cgggaccatg taccaacacg gagagatctt cagtgcccat gagctgttcc 780
cctcccgcct gcccaaccag tgtgtcctct gcagctgcac agagggccag atctactgcg 840
gcctcacaac ctgccccgaa ccaggctgcc cagcaccccc cccgctgcca gactcctgct 900
gccaagcctg caaagatgag gcaagtgagc aatcggatga agaggacagt gtgcagtcgc 960
tccatggggt gagacatcct caggatccat gttccagtga tgctgggaga aagagaggcc 1020
cgggcacccc agcccccact ggcctcagcg cccctctgag cttcatccct cgccacttca 1080
gacccaaggg agcaggcagc acaactgtca agatcgtcct gaaggagaaa cataagaaag 1140
aggacaaagc agaccctggc cacagtgaga tcagttctac caggtgtccc aaggcaccgg 1200
gccgggtcct cgtccacaca tcggtatccc caagcccaga caacctgcgt cgctttgccc 1260
tggaacacga ggcctcggac ttggtggaga tctacctctg gaagctggta aaagatgagg 1320
aaactgaggc tcagagaggt gaagtacctg gcccaaggcc acacagccag aatcttccac 1380
ttgactcaga tcaagaaagt caggaagcaa gacttccaga aagaggcaca gcacttccga 1440
ctgctcgctg gcccccacga aggtcactgg aacgtcttcc tagcccagac cctggagctg 1500
aaggtcacgg ccagtccaga caaagtgacc aagacataac aaagacctaa cagttgcaga 1560
tatgagctgt ataattgttg ttattatata ttaataaata agaagttgca taaccatcaa 1620
aa 1622

<210> 10
<211> 1567
<212> DNA
<213> Homo sapiens

<400> 10
ggacaaataa aaaggaaaca agcatgattg tgagggcaga ggagcgtggg actgagtcag 60
gagactggtg ctgtcatcgc tgcctggtga ctgacttgct gtgtggccct caggtgtaac 120
ttaccctctc tgggcctcat ttgtctaatc ataataatta acgctgatac catgatataa 180
atctgtacag catttcactg cttgattccc taactgccct gtgagataag cgttaaggct 240
cagagacagt ggcatgccca gtgttgcaca gtaagtgtgt ggtaaagccg agattcaaac 300
tcagaccttc tggccccttg cctaggagag catgcccagt tgtctagcag attctctttt 360
gcctgagtgg cccagatgac atctctttta gagctagaaa gaaggagaaa tgagacaggg 420
tctttgggct ggagcctcct gggactaaca tggcactggt cggtttgcca ggcccagaca 480
tgttctgcct tttccatggg aagagatact cccccggcga gagctggcac ccctacttgg 540
agccacaagg cctgatgtac tgcctgcgct gtacctgctc agagggcgcc catgtgagtt 600
gttaccgcct ccactgtccg cctgtccact gcccccagcc tgtgacggag ccacagcaat 660
gctgtcccaa gtgtgtggaa cctcacactc cctctggact ccgggcccca ccaaagtcct 720
gccagcacaa cgggaccatg taccaacacg gagagatctt cagtgcccat gagctgttcc 780
cctcccgcct gcccaaccag tgtgtcctct gcagctgcac agagggccag atctactgcg 840
gcctcacaac ctgccccgaa ccaggctgcc cagcaccccc cccgctgcca gactcctgct 900

gccaagcctg caaagatgag gcaagtgagc aatcggatga agaggacagt gtgcagtcgc 960
tccatggggt gagacatcct caggatccat gttccagtga tgctgggaga aagagaggcc 1020
cgggcacccc agcccccact ggcctcagcg cccctctgag cttcatccct cgccacttca 1080
gacccaaggg agcaggcagc acaactgtca agatcgtcct gaaggagaaa cataagaaag 1140
aggacaaagc agaccctggc cacagtgaga tcagttctac caggtgtccc aaggcaccgg 1200
gccgggtcct cgtccacaca tcggtatccc caagcccaga caacctgcgt cgctttgccc 1260
tggaacacga ggcctcggac ttggtggaga tctacctctg gaagctggta aaaggaatct 1320
tccacttgac tcagatcaag aaagtcagga agcaagactt ccagaaagag gcacagcact 1380
tccgactgct cgctggcccc cacgaaggtc actggaacgt cttcctagcc cagaccctgg 1440
agctgaaggt cacggccagt ccagacaaag tgaccaagac ataacaaaga cctaacagtt 1500
gcagatatga gctgtataat tgttgttatt atatattaat aaataagaag ttgcataacc 1560
atcaaaa 1567

<210> 11
<211> 1202
<212> DNA
<213> Mouse

<400> 11
atttctctat tcctgatccc acactgctct gcctacccac accagcccca aggtctnaga 60
aagccctgga ggctggcttg ccaaatcctt gtcagtgtnt ttattgatta gtctgagaat 120
atcttagacc tcacccacaa ggttctgtgt ggagcctgtg ctctctgtct gtctgtctgt 180
ctgtctgtct gtctgtctgt ctgcctgcct ctctctgtct gtctccgtct gtctctgtct 240
ctctgtctgt ctctgtctgt ctctttctct ctgtctctct ctgtgtctct gtctctgtct 300
ctgtctctct ctctctctca gaagtcctct agccttctct agcaggcgtc tcatgcagcc 360
tggttggtgt tcccagctgt ggcctatccc acagacagct ccacatcctg cttgctgttc 420
gcagagacat tcccaggatc catgctcgga gaggagaggc cccagcacgc cagcccccac 480
cagcctcagc tcccctctgg gcttcatccn tcgccacttc cagtcagtag gaatgggcag 540
cacaaccatc aagattatct tgaaggagaa acataaaaaa gcttgcacac acaatgggaa 600
gacatactcc catggggagg tgtggcaccc cactgtgctc tcctttggcc ccatgccctg 660
catcctgtgc acatgtattg atggctacca ggactgccac cgtgtgacct gccccaccca 720
atatccctgc agtcaaccca agaaagtggc tgggaagtgc tgcaagatct gcccagagga 780
cgaggcggaa gatgaccaca gtgaggtcat ttccacccgg tgtcccaagg taccaggcca 840
gttccaggtg tacacgttgg catctccaag cccagacagc ctacaccgct ttgtcctgga 900
gcatgaagcc tctgaccagg tagagatgta catttggaag ctggtgaaag gaatttacca 960
cttggttcag atcaagagag tcaggaagca agatttccag aaagaggttc agaacttccg 1020
gctgctcacc ggcacccatg aaggttactg gaccgttttc ctagcccaga ttccagagct 1080
gaaagttaca gccagcccag acaaagtgac caagacatta tagcaaggac ctaaagagtt 1140
gcagatacga gttttattgg ttttgttatt atatattaat aaagaagtcg cattaccctt 1200
tc 1202

<210> 12
<211> 398
<212> PRT
<213> Homo sapiens

<400> 12
Arg Glu Pro Gln Gly Leu Met Tyr Cys Leu Arg Cys Thr Cys Ser Glu
1 5 10 15

Gly Ala His Val Ser Cys Tyr Arg Leu His Cys Pro Pro Val His Cys
20 25 30

Pro Gln Pro Val Thr Glu Pro Gln Gln Cys Cys Pro Lys Cys Val Glu
35 40 45

Pro His Thr Pro Ser Gly Leu Arg Ala Pro Pro Lys Ser Cys Gln His
50 55 60

Asn Gly Thr Met Tyr Gln His Gly Glu Ile Phe Ser Ala His Glu Leu
65 70 75 80

Phe Pro Ser Arg Leu Pro Asn Gln Cys Val Leu Cys Ser Cys Thr Glu
85 90 95

Gly Gln Ile Tyr Cys Gly Leu Thr Thr Cys Pro Glu Pro Gly Cys Pro
100 105 110

Ala Pro Leu Pro Leu Pro Asp Ser Cys Cys Gln Ala Cys Lys Asp Glu
115 120 125

Ala Ser Glu Gln Ser Asp Glu Glu Asp Ser Val Gln Ser Leu His Gly
130 135 140

Val Arg His Pro Gln Asp Pro Cys Ser Ser Asp Ala Gly Arg Lys Arg
145 150 155 160

Gly Pro Gly Thr Pro Ala Pro Thr Gly Leu Ser Ala Pro Leu Ser Phe
165 170 175

Ile Pro Arg His Phe Arg Pro Lys Gly Ala Gly Ser Thr Thr Val Lys
180 185 190

Ile Val Leu Lys Glu Lys His Xaa Lys Ala Cys Val His Gly Gly Lys
195 200 205

Thr Tyr Ser His Gly Glu Val Trp His Pro Ala Phe Arg Ala Phe Gly
210 215 220

Pro Cys Pro Cys Ile Leu Cys Thr Cys Glu Asp Gly Arg Gln Asp Cys
225 230 235 240

Gln Arg Val Thr Cys Pro Thr Lys Tyr Pro Cys Arg His Pro Glu Lys
245 250 255

Val Ala Gly Lys Cys Cys Lys Ile Cys Pro Glu Asp Lys Ala Asp Pro
260 265 270

Gly His Ser Glu Ile Ser Ser Thr Arg Cys Pro Lys Ala Pro Gly Arg
275 280 285

Val Leu Val His Thr Ser Val Ser Pro Ser Pro Asp Asn Leu Arg Arg
290 295 300

Phe Ala Leu Glu His Glu Ala Ser Asp Leu Val Glu Ile Tyr Leu Trp
305 310 315 320

Lys Leu Val Lys Asp Glu Glu Thr Glu Ala Gln Arg Gly Glu Val Pro
325 330 335

Gly Pro Arg Pro His Ser Gln Asn Phe His Leu Thr Gln Ile Lys Lys
340 345 350

Val Arg Lys Gln Asp Phe Gln Lys Glu Ala Gln His Phe Arg Leu Leu
355 360 365

Ala Gly Pro His Glu Gly His Trp Asn Val Phe Leu Ala Gln Thr Leu
370 375 380

Glu Leu Lys Val Thr Ala Ser Pro Asp Lys Val Thr Lys Thr

385 390 395

<210> 13
<211> 539
<212> PRT
<213> Homo sapiens

<400> 13
Ser Pro Leu Pro Ser Ala Gly Pro Ser Phe Val Ser Pro Ser Leu Pro
1 5 10 15

Pro Phe Pro Ala Phe Ser Phe His Leu Ser Leu Leu Pro Thr Leu Asp
20 25 30

Leu Pro Ser Cys Pro Pro Phe Leu Pro Thr Ala Ala Ser Trp Pro Phe
35 40 45

Ser Asp Pro Ala Leu Ala Ala Asp Leu Leu Gly Ser Cys Gly Leu Ile
50 55 60

Cys Gly Pro Cys Xaa Ser Val Ser Phe Ser Ser Pro Val Leu Pro Thr
65 70 75 80

Pro Leu Pro Asp Gln Arg Pro Asp Pro Gly Glu Arg Met Val Pro Glu
85 90 95

Val Arg Val Leu Ser Ser Leu Leu Gly Leu Ala Leu Leu Trp Phe Pro
100 105 110

Leu Asp Ser His Ala Arg Ala Arg Pro Asp Met Phe Cys Leu Phe His
115 120 125

Gly Lys Arg Tyr Ser Pro Gly Glu Ser Trp His Pro Tyr Leu Glu Pro
130 135 140

Gln Gly Leu Met Tyr Cys Leu Arg Cys Thr Cys Ser Glu Gly Ala His
145 150 155 160

Val Ser Cys Tyr Arg Leu His Cys Pro Pro Val His Cys Pro Gln Pro
165 170 175

Val Thr Glu Pro Gln Gln Cys Cys Pro Lys Cys Val Glu Pro His Thr
180 185 190

Pro Ser Gly Leu Arg Ala Pro Pro Lys Ser Cys Gln His Asn Gly Thr
195 200 205

Met Tyr Gln His Gly Glu Ile Phe Ser Ala His Glu Leu Phe Pro Ser
210 215 220

Arg Leu Pro Asn Gln Cys Val Leu Cys Ser Cys Thr Glu Gly Gln Ile
225 230 235 240

Tyr Cys Gly Leu Thr Thr Cys Pro Glu Pro Gly Cys Pro Ala Pro Leu
245 250 255

Pro Leu Pro Asp Ser Cys Cys Gln Ala Cys Lys Asp Glu Ala Ser Glu
260 265 270

Gln Ser Asp Glu Glu Asp Ser Val Gln Ser Leu His Gly Val Arg His

275 280 285

Pro Gln Asp Pro Cys Ser Ser Asp Ala Gly Arg Lys Arg Gly Pro Gly
290 295 300

Thr Pro Ala Pro Thr Gly Leu Ser Ala Pro Leu Ser Phe Ile Pro Arg
305 310 315 320

His Phe Arg Pro Lys Gly Ala Gly Ser Thr Thr Val Lys Ile Val Leu
325 330 335

Lys Glu Lys His Xaa Lys Ala Cys Val His Gly Gly Lys Thr Tyr Ser
340 345 350

His Gly Glu Val Trp His Pro Ala Phe Arg Ala Phe Gly Pro Cys Pro
355 360 365

Cys Ile Leu Cys Thr Cys Glu Asp Gly Arg Gln Asp Cys Gln Arg Val
370 375 380

Thr Cys Pro Thr Lys Tyr Pro Cys Arg His Pro Glu Lys Val Ala Gly
385 390 395 400

Lys Cys Cys Lys Ile Cys Pro Glu Asp Lys Ala Asp Pro Gly His Ser
405 410 415

Glu Ile Ser Ser Thr Arg Cys Pro Lys Ala Pro Gly Arg Val Leu Val
420 425 430

His Thr Ser Val Ser Pro Ser Pro Asp Asn Leu Arg Arg Phe Ala Leu
435 440 445

Glu His Glu Ala Ser Asp Leu Val Glu Ile Tyr Leu Trp Lys Leu Val
450 455 460

Lys Asp Glu Glu Thr Glu Ala Gln Arg Gly Glu Val Pro Gly Pro Arg
465 470 475 480

Pro His Ser Gln Asn Phe His Leu Thr Gln Ile Lys Lys Val Arg Lys
485 490 495

Gln Asp Phe Gln Lys Glu Ala Gln His Phe Arg Leu Leu Ala Gly Pro
500 505 510

His Glu Gly His Trp Asn Val Phe Leu Ala Gln Thr Leu Glu Leu Lys
515 520 525

Val Thr Ala Ser Pro Asp Lys Val Thr Lys Thr
530 535

<210> 14
<211> 388
<212> PRT
<213> Homo sapiens

<400> 14
Ile Ser Ser Trp Gly Gln Met Gln Asn His Gln Lys Ser Gly Leu Val
1 5 10 15

Asn Phe Ser Lys Asp Ser His Glu Thr Ser Phe Ser Ser Ser Ser Cys

20 25 30

Pro Ser Pro Thr Val Glu Pro His Thr Pro Ser Gly Leu Arg Ala Pro
35 40 45

Pro Lys Ser Cys Gln His Asn Gly Thr Met Tyr Gln His Gly Glu Ile
50 55 60

Phe Ser Ala His Glu Leu Phe Pro Ser Arg Leu Pro Asn Gln Cys Val
65 70 75 80

Leu Cys Ser Cys Thr Glu Gly Gln Ile Tyr Cys Gly Leu Thr Thr Cys
85 90 95

Pro Glu Pro Gly Cys Pro Ala Pro Leu Pro Leu Pro Asp Ser Cys Cys
100 105 110

Gln Ala Cys Lys Asp Glu Ala Ser Glu Gln Ser Asp Glu Glu Asp Ser
115 120 125

Val Gln Ser Leu His Gly Val Arg His Pro Gln Asp Pro Cys Ser Ser
130 135 140

Asp Ala Gly Arg Lys Arg Gly Pro Gly Thr Pro Ala Pro Thr Gly Leu
145 150 155 160

Ser Ala Pro Leu Ser Phe Ile Pro Arg His Phe Arg Pro Lys Gly Ala
165 170 175

Gly Ser Thr Thr Val Lys Ile Val Leu Lys Glu Lys His Xaa Lys Ala
180 185 190

Cys Val His Gly Gly Lys Thr Tyr Ser His Gly Glu Val Trp His Pro
195 200 205

Ala Phe Arg Ala Phe Gly Pro Cys Pro Cys Ile Leu Cys Thr Cys Glu
210 215 220

Asp Gly Arg Gln Asp Cys Gln Arg Val Thr Cys Pro Thr Lys Tyr Pro
225 230 235 240

Cys Arg His Pro Glu Lys Val Ala Gly Lys Cys Cys Lys Ile Cys Pro
245 250 255

Glu Asp Lys Ala Asp Pro Gly His Ser Glu Ile Ser Ser Thr Arg Cys
260 265 270

Pro Lys Ala Pro Gly Arg Val Leu Val His Thr Ser Val Ser Pro Ser
275 280 285

Pro Asp Asn Leu Arg Arg Phe Ala Leu Glu His Glu Ala Ser Asp Leu
290 295 300

Val Glu Ile Tyr Leu Trp Lys Leu Val Lys Asp Glu Glu Thr Glu Ala
305 310 315 320

Gln Arg Gly Glu Val Pro Gly Pro Arg Pro His Ser Gln Asn Phe His
325 330 335

Leu Thr Gln Ile Lys Lys Val Arg Lys Gln Asp Phe Gln Lys Glu Ala
340 345 350

Gln His Phe Arg Leu Leu Ala Gly Pro His Glu Gly His Trp Asn Val
355 360 365

Phe Leu Ala Gln Thr Leu Glu Leu Lys Val Thr Ala Ser Pro Asp Lys
370 375 380

Val Thr Lys Thr
385

<210> 15
<211> 439
<212> PRT
<213> Homo sapiens

<400> 15
Asp Arg Val Phe Gly Leu Glu Pro Pro Gly Thr Asn Met Ala Leu Val
1 5 10 15

Gly Leu Pro Gly Pro Asp Met Phe Cys Leu Phe His Gly Lys Arg Tyr
20 25 30

Ser Pro Gly Glu Ser Trp His Pro Tyr Leu Glu Pro Gln Gly Leu Met
35 40 45

Tyr Cys Leu Arg Cys Thr Cys Ser Glu Gly Ala His Val Ser Cys Tyr
50 55 60

Arg Leu His Cys Pro Pro Val His Cys Pro Gln Pro Val Thr Glu Pro
65 70 75 80

Gln Gln Cys Cys Pro Lys Cys Val Glu Pro His Thr Pro Ser Gly Leu
85 90 95

Arg Ala Pro Pro Lys Ser Cys Gln His Asn Gly Thr Met Tyr Gln His
100 105 110

Gly Glu Ile Phe Ser Ala His Glu Leu Phe Pro Ser Arg Leu Pro Asn
115 120 125

Gln Cys Val Leu Cys Ser Cys Thr Glu Gly Gln Ile Tyr Cys Gly Leu
130 135 140

Thr Thr Cys Pro Glu Pro Gly Cys Pro Ala Pro Leu Pro Leu Pro Asp
145 150 155 160

Ser Cys Cys Gln Ala Cys Lys Asp Glu Ala Ser Glu Gln Ser Asp Glu
165 170 175

Glu Asp Ser Val Gln Ser Leu His Gly Val Arg His Pro Gln Asp Pro
180 185 190

Cys Ser Ser Asp Ala Gly Arg Lys Arg Gly Pro Gly Thr Pro Ala Pro
195 200 205

Thr Gly Leu Ser Ala Pro Leu Ser Phe Ile Pro Arg His Phe Arg Pro
210 215 220

Lys Gly Ala Gly Ser Thr Thr Val Lys Ile Val Leu Lys Glu Lys His
225 230 235 240

Xaa Lys Ala Cys Val His Gly Gly Lys Thr Tyr Ser His Gly Glu Val
245 250 255

Trp His Pro Ala Phe Arg Ala Phe Gly Pro Cys Pro Cys Ile Leu Cys
260 265 270

Thr Cys Glu Asp Gly Arg Gln Asp Cys Gln Arg Val Thr Cys Pro Thr
275 280 285

Lys Tyr Pro Cys Arg His Pro Glu Lys Val Ala Gly Lys Cys Cys Lys
290 295 300

Ile Cys Pro Glu Asp Lys Ala Asp Pro Gly His Ser Glu Ile Ser Ser
305 310 315 320

Thr Arg Cys Pro Lys Ala Pro Gly Arg Val Leu Val His Thr Ser Val
325 330 335

Ser Pro Ser Pro Asp Asn Leu Arg Arg Phe Ala Leu Glu His Glu Ala
340 345 350

Ser Asp Leu Val Glu Ile Tyr Leu Trp Lys Leu Val Lys Asp Glu Glu
355 360 365

Thr Glu Ala Gln Arg Gly Glu Val Pro Gly Pro Arg Pro His Ser Gln
370 375 380

Asn Phe His Leu Thr Gln Ile Lys Lys Val Arg Lys Gln Asp Phe Gln
385 390 395 400

Lys Glu Ala Gln His Phe Arg Leu Leu Ala Gly Pro His Glu Gly His
405 410 415

Trp Asn Val Phe Leu Ala Gln Thr Leu Glu Leu Lys Val Thr Ala Ser
420 425 430

Pro Asp Lys Val Thr Lys Thr
435

<210> 16
<211> 549
<212> PRT
<213> Homo sapiens

<400> 16
Thr Phe Pro Leu Ser Leu Ile Ala Ser Pro Phe Cys Trp Thr Phe Leu
1 5 10 15

Arg Leu Ser Ile Ser Pro Ser Phe Pro Arg Val Leu Phe Pro Pro Phe
20 25 30

Ser Ser Ser His Leu Arg Pro Pro Phe Leu Pro Ser Phe Pro Ala His
35 40 45

Arg Cys Phe Leu Ala Leu Leu Arg Pro Arg Ser Ser Ser Arg Pro Pro
50 55 60

Gly Val Cys Gly Leu Ile Cys Gly Pro Cys Ala Ser Val Ser Phe Ser
65 70 75 80

Ser Pro Phe Leu Pro Thr Pro Leu Pro Asp Gln Arg Pro Asp Pro Gly
85 90 95

Glu Arg Met Val Pro Glu Val Arg Val Leu Ser Ser Leu Leu Gly Leu
100 105 110

Ala Leu Leu Trp Phe Pro Leu Asp Ser His Ala Arg Ala Arg Pro Asp
115 120 125

Met Phe Cys Leu Phe His Gly Lys Arg Tyr Ser Pro Gly Glu Ser Trp
130 135 140

His Pro Tyr Leu Glu Pro Gln Gly Leu Met Tyr Cys Leu Arg Cys Thr
145 150 155 160

Cys Ser Glu Gly Ala His Val Ser Cys Tyr Arg Leu His Cys Pro Pro
165 170 175

Val His Cys Pro Gln Pro Val Thr Glu Pro Gln Gln Cys Cys Pro Lys
180 185 190

Cys Val Glu Pro His Thr Pro Ser Gly Leu Arg Ala Pro Pro Lys Ser
195 200 205

Cys Gln His Asn Gly Thr Met Tyr Gln His Gly Glu Ile Phe Ser Ala
210 215 220

His Glu Leu Phe Pro Ser Arg Leu Pro Asn Gln Cys Val Leu Cys Ser
225 230 235 240

Cys Thr Glu Gly Gln Ile Tyr Cys Gly Leu Thr Thr Cys Pro Glu Pro
245 250 255

Gly Cys Pro Ala Pro Leu Pro Leu Pro Asp Ser Cys Cys Gln Ala Cys
260 265 270

Lys Asp Glu Ala Ser Glu Gln Ser Asp Glu Glu Asp Arg Val Gln Ser
275 280 285

Leu His Gly Val Arg His Pro Gln Asp Pro Cys Ser Ser Asp Ala Gly
290 295 300

Arg Lys Arg Gly Pro Gly Thr Pro Ala Pro Thr Gly Leu Ser Ala Pro
305 310 315 320

Leu Ser Phe Ile Pro Arg His Phe Ile Pro Lys Gly Ala Gly Ser Thr
325 330 335

Thr Val Lys Ile Val Leu Lys Glu Lys His Lys Lys Ala Cys Val His
340 345 350

Gly Gly Lys Thr Tyr Ser His Gly Glu Val Trp His Pro Ala Phe Arg
355 360 365

Ala Phe Gly Pro Leu Pro Cys Ile Leu Cys Thr Cys Glu Asp Gly Arg
370 375 380

Gln Asp Cys Gln Arg Val Thr Cys Pro Thr Glu Tyr Pro Cys Arg His
385 390 395 400

Pro Glu Lys Val Ala Gly Lys Cys Cys Lys Ile Cys Pro Glu Asp Lys
405 410 415

Ala Asp Pro Gly His Ser Glu Ile Ser Ser Thr Arg Cys Pro Lys Ala
420 425 430

Pro Gly Arg Val Leu Val His Thr Ser Val Ser Pro Ser Pro Asp Asn
435 440 445

Leu Arg Arg Phe Ala Leu Glu His Glu Ala Ser Asp Leu Val Glu Ile
450 455 460

Tyr Leu Trp Lys Leu Val Lys Asp Glu Glu Thr Glu Ala Gln Arg Gly
465 470 475 480

Glu Val Pro Gly Pro Arg Pro His Ser Gln Asn Leu Pro Leu Asp Ser
485 490 495

Asp Gln Glu Ser Gln Glu Ala Arg Leu Pro Glu Arg Gly Thr Ala Leu
500 505 510

Pro Thr Ala Arg Trp Pro Pro Arg Arg Ser Leu Glu Arg Leu Pro Ser
515 520 525

Pro Asp Pro Gly Ala Glu Gly His Gly Gln Ser Arg Gln Ser Asp Gln
530 535 540

Asp Ile Thr Lys Thr
545

<210> 17
<211> 549
<212> PRT
<213> Homo sapiens

<400> 17
Thr Phe Pro Leu Ser Leu Ile Ala Ser Pro Phe Cys Trp Thr Phe Leu
1 5 10 15

Arg Leu Ser Ile Ser Pro Ser Phe Pro Arg Val Leu Phe Pro Pro Phe
20 25 30

Ser Ser Ser His Leu Arg Pro Pro Phe Leu Pro Ser Phe Pro Ala His
35 40 45

Arg Cys Phe Leu Ala Leu Leu Arg Pro Arg Ser Ser Ser Arg Pro Pro
50 55 60

Gly Val Cys Gly Leu Ile Cys Gly Pro Cys Ala Ser Val Ser Phe Ser
65 70 75 80

Ser Pro Phe Leu Pro Thr Pro Leu Pro Asp Gln Arg Pro Asp Pro Gly
85 90 95

Glu Arg Met Val Pro Glu Val Arg Val Leu Ser Ser Leu Leu Gly Leu
100 105 110

Ala Leu Leu Trp Phe Pro Leu Asp Ser His Ala Arg Ala Arg Pro Asp
115 120 125

Met Phe Cys Leu Phe His Gly Lys Arg Tyr Ser Pro Gly Glu Ser Trp
130 135 140

His Pro Tyr Leu Glu Pro Gln Gly Leu Met Tyr Cys Leu Arg Cys Thr
145 150 155 160

Cys Ser Glu Gly Ala His Val Ser Cys Tyr Arg Leu His Cys Pro Pro
165 170 175

Val His Cys Pro Gln Pro Val Thr Glu Pro Gln Gln Cys Cys Pro Lys
180 185 190

Cys Val Glu Pro His Thr Pro Ser Gly Leu Arg Ala Pro Pro Lys Ser
195 200 205

Cys Gln His Asn Gly Thr Met Tyr Gln His Gly Glu Ile Phe Ser Ala
210 215 220

His Glu Leu Phe Pro Ser Arg Leu Pro Asn Gln Cys Val Leu Cys Ser
225 230 235 240

Cys Thr Glu Gly Gln Ile Tyr Cys Gly Leu Thr Thr Cys Pro Glu Pro
245 250 255

Gly Cys Pro Ala Pro Leu Pro Leu Pro Asp Ser Cys Cys Gln Ala Cys
260 265 270

Lys Gly Glu Ala Ser Glu Gln Ser Asp Glu Glu Asp Ser Val Gln Ser
275 280 285

Leu His Gly Val Arg His Pro Gln Asp Pro Cys Ser Ser Asp Ala Gly
290 295 300

Arg Lys Arg Gly Pro Gly Thr Pro Ala Pro Thr Gly Leu Ser Ala Pro
305 310 315 320

Leu Ser Phe Ile Pro Arg His Phe Arg Pro Lys Gly Ala Gly Ser Thr
325 330 335

Thr Val Lys Ile Val Leu Lys Glu Lys His Lys Lys Ala Cys Val His
340 345 350

Gly Gly Lys Thr Tyr Ser His Gly Glu Val Trp His Pro Ala Phe Arg
355 360 365

Ala Phe Gly Pro Leu Pro Cys Ile Leu Cys Thr Cys Glu Asp Gly Arg
370 375 380

Gln Asp Cys Gln Arg Val Thr Cys Pro Thr Glu Tyr Pro Cys Arg His
385 390 395 400

Pro Glu Lys Val Ala Gly Lys Cys Cys Lys Ile Cys Pro Glu Asp Lys
405 410 415

Ala Asp Pro Gly His Ser Glu Ile Ser Ser Thr Arg Cys Pro Lys Ala
420 425 430

Pro Gly Arg Val Leu Val His Thr Ser Val Ser Pro Ser Pro Asp Asn
435 440 445

Leu Arg Arg Phe Ala Leu Glu His Glu Ala Ser Asp Leu Val Glu Ile

450 455 460

Tyr Leu Trp Lys Leu Val Lys Asp Glu Glu Thr Glu Ala Gln Arg Gly
465 470 475 480

Glu Val Pro Gly Pro Arg Pro His Ser Gln Asn Leu Pro Leu Asp Ser
485 490 495

Asp Gln Glu Ser Gln Glu Ala Arg Leu Pro Glu Arg Gly Thr Ala Leu
500 505 510

Pro Thr Ala Arg Trp Pro Pro Arg Arg Ser Leu Glu Arg Leu Pro Ser
515 520 525

Pro Asp Pro Gly Ala Glu Gly His Gly Gln Ser Arg Gln Ser Asp Gln
530 535 540

Asp Ile Thr Lys Thr
545

<210> 18
<211> 392
<212> PRT
<213> Homo sapiens

<400> 18
Ile Ser Ser Trp Gly Gln Met Gln Asn His Gln Lys Ser Gly Leu Val
1 5 10 15

Asn Phe Ser Lys Asp Ser His Glu Thr Ser Phe Ser Ser Ser Ser Cys
20 25 30

Pro Ser Pro Thr Ala Glu Pro His Thr Pro Ser Gly Leu Arg Ala Pro
35 40 45

Pro Lys Ser Cys Gln His Asn Gly Thr Met Tyr Gln His Gly Glu Ile
50 55 60

Phe Ser Ala His Glu Leu Phe Pro Ser Arg Leu Pro Asn Gln Cys Val
65 70 75 80

Leu Cys Ser Cys Thr Glu Gly Gln Ile Tyr Cys Gly Leu Thr Thr Cys
85 90 95

Pro Glu Pro Gly Cys Pro Ala Pro Leu Pro Leu Pro Asp Ser Cys Cys
100 105 110

Gln Ala Cys Lys Asp Glu Ala Ser Glu Gln Ser Asp Glu Glu Asp Ser
115 120 125

Val Gln Ser Leu His Gly Val Arg His Pro Gln Asp Pro Cys Ser Ser
130 135 140

Asp Ala Gly Arg Lys Arg Gly Pro Gly Thr Pro Ala Pro Thr Gly Leu
145 150 155 160

Ser Ala Pro Leu Ser Phe Ile Pro Arg His Phe Arg Pro Lys Gly Ala
165 170 175

Gly Ser Thr Thr Val Lys Ile Val Leu Lys Glu Lys His Lys Lys Ala

180 185 190

Cys Val His Gly Gly Lys Thr Tyr Ser His Gly Glu Val Trp His Pro
195 200 205

Ala Phe Arg Ala Phe Gly Pro Leu Pro Cys Ile Leu Cys Thr Cys Glu
210 215 220

Asp Gly Arg Gln Asp Cys Gln Arg Val Thr Cys Pro Thr Glu Tyr Pro
225 230 235 240

Cys Arg His Pro Glu Lys Val Ala Gly Lys Cys Cys Lys Ile Cys Pro
245 250 255

Glu Asp Lys Ala Asp Pro Gly His Ser Glu Ile Ser Ser Thr Arg Cys
260 265 270

Pro Lys Ala Pro Gly Arg Val Leu Val His Thr Ser Val Ser Pro Ser
275 280 285

Pro Asp Asn Leu Arg Arg Phe Ala Leu Glu His Glu Ala Ser Asp Leu
290 295 300

Val Glu Ile Tyr Leu Trp Lys Leu Val Lys Asp Glu Glu Thr Glu Ala
305 310 315 320

Gln Arg Gly Glu Val Pro Gly Pro Arg Pro His Ser Gln Asn Leu Pro
325 330 335

Leu Asp Ser Asp Gln Glu Ser Gln Glu Ala Arg Leu Pro Glu Arg Gly
340 345 350

Thr Ala Leu Pro Thr Ala Arg Trp Pro Pro Arg Arg Ser Leu Glu Arg
355 360 365

Leu Pro Ser Pro Asp Pro Gly Ala Glu Gly His Gly Gln Ser Arg Gln
370 375 380

Ser Asp Gln Asp Ile Thr Lys Thr
385 390

<210> 19
<211> 443
<212> PRT
<213> Homo sapiens

<400> 19
Asp Arg Val Phe Gly Leu Glu Pro Pro Gly Thr Asn Met Ala Leu Val
1 5 10 15

Gly Leu Pro Gly Pro Asp Met Phe Cys Leu Phe His Gly Lys Arg Tyr
20 25 30

Ser Pro Gly Glu Ser Trp His Pro Tyr Leu Glu Pro Gln Gly Leu Met
35 40 45

Tyr Cys Leu Arg Cys Thr Cys Ser Glu Gly Ala His Val Ser Cys Tyr
50 55 60

Arg Leu His Cys Pro Pro Val His Cys Pro Gln Pro Val Thr Glu Pro

65 70 75 80

Gln Gln Cys Cys Pro Lys Cys Val Glu Pro His Thr Pro Ser Gly Leu
85 90 95

Arg Ala Pro Pro Lys Ser Cys Gln His Asn Gly Thr Met Tyr Gln His
100 105 110

Gly Glu Ile Phe Ser Ala His Glu Leu Phe Pro Ser Arg Leu Pro Asn
115 120 125

Gln Cys Val Leu Cys Ser Cys Thr Glu Gly Gln Ile Tyr Cys Gly Leu
130 135 140

Thr Thr Cys Pro Glu Pro Gly Cys Pro Ala Pro Leu Pro Leu Pro Asp
145 150 155 160

Ser Cys Cys Gln Ala Cys Lys Asp Glu Ala Ser Glu Gln Ser Asp Glu
165 170 175

Glu Asp Ser Val Gln Ser Leu His Gly Val Arg His Pro Gln Asp Pro
180 185 190

Cys Ser Ser Asp Ala Gly Arg Lys Arg Gly Pro Gly Thr Pro Ala Pro
195 200 205

Thr Gly Leu Ser Ala Pro Leu Ser Phe Ile Pro Arg His Phe Arg Pro
210 215 220

Lys Gly Ala Gly Ser Thr Thr Val Lys Ile Val Leu Lys Glu Lys His
225 230 235 240

Lys Lys Ala Cys Val His Gly Gly Lys Thr Tyr Ser His Gly Glu Val
245 250 255

Trp His Pro Ala Phe Arg Ala Phe Gly Pro Leu Pro Cys Ile Leu Cys
260 265 270

Thr Cys Glu Asp Gly Arg Gln Asp Cys Gln Arg Val Thr Cys Pro Thr
275 280 285

Glu Tyr Pro Cys Arg His Pro Glu Lys Val Ala Gly Lys Cys Cys Lys
290 295 300

Ile Cys Pro Glu Asp Lys Ala Asp Pro Gly His Ser Glu Ile Ser Ser
305 310 315 320

Thr Arg Cys Pro Lys Ala Pro Gly Arg Val Leu Val His Thr Ser Val
325 330 335

Ser Pro Ser Pro Asp Asn Leu Arg Arg Phe Ala Leu Glu His Glu Ala
340 345 350

Ser Asp Leu Val Glu Ile Tyr Leu Trp Lys Leu Val Lys Asp Glu Glu
355 360 365

Thr Glu Ala Gln Arg Gly Glu Val Pro Gly Pro Arg Pro His Ser Gln
370 375 380

Asn Leu Pro Leu Asp Ser Asp Gln Glu Ser Gln Glu Ala Arg Leu Pro
385 390 395 400

Glu Arg Gly Thr Ala Leu Pro Thr Ala Arg Trp Pro Pro Arg Arg Ser
405 410 415

Leu Glu Arg Leu Pro Ser Pro Asp Pro Gly Ala Glu Gly His Gly Gln
420 425 430

Ser Arg Gln Ser Asp Gln Asp Ile Thr Lys Thr
435 440

<210> 20
<211> 378
<212> PRT
<213> Homo sapiens

<400> 20
Asp Arg Val Phe Gly Leu Glu Pro Pro Gly Thr Asn Met Ala Leu Val
1 5 10 15

Gly Leu Pro Gly Pro Asp Met Phe Cys Leu Phe His Gly Lys Arg Tyr
20 25 30

Ser Pro Gly Glu Ser Trp His Pro Tyr Leu Glu Pro Gln Gly Leu Met
35 40 45

Tyr Cys Leu Arg Cys Thr Cys Ser Glu Gly Ala His Val Ser Cys Tyr
50 55 60

Arg Leu His Cys Pro Pro Val His Cys Pro Gln Pro Val Thr Glu Pro
65 70 75 80

Gln Gln Cys Cys Pro Lys Cys Val Glu Pro His Thr Pro Ser Gly Leu
85 90 95

Arg Ala Pro Pro Lys Ser Cys Gln His Asn Gly Thr Met Tyr Gln His
100 105 110

Gly Glu Ile Phe Ser Ala His Glu Leu Phe Pro Ser Arg Leu Pro Asn
115 120 125

Gln Cys Val Leu Cys Ser Cys Thr Glu Gly Gln Ile Tyr Cys Gly Leu
130 135 140

Thr Thr Cys Pro Glu Pro Gly Cys Pro Ala Pro Leu Pro Leu Pro Asp
145 150 155 160

Ser Cys Cys Gln Ala Cys Lys Asp Glu Ala Ser Glu Gln Ser Asp Glu
165 170 175

Glu Asp Ser Val Gln Ser Leu His Gly Val Arg His Pro Gln Asp Pro
180 185 190

Cys Ser Ser Asp Ala Gly Arg Lys Arg Gly Pro Gly Thr Pro Ala Pro
195 200 205

Thr Gly Leu Ser Ala Pro Leu Ser Phe Ile Pro Arg His Phe Arg Pro
210 215 220

Lys Gly Ala Gly Ser Thr Thr Val Lys Ile Val Leu Lys Glu Lys His
225 230 235 240

Lys Lys Glu Asp Lys Ala Asp Pro Gly His Ser Glu Ile Ser Ser Thr
245 250 255

Arg Cys Pro Lys Ala Pro Gly Arg Val Leu Val His Thr Ser Val Ser
260 265 270

Pro Ser Pro Asp Asn Leu Arg Arg Phe Ala Leu Glu His Glu Ala Ser
275 280 285

Asp Leu Val Glu Ile Tyr Leu Trp Lys Leu Val Lys Asp Glu Glu Thr
290 295 300

Glu Ala Gln Arg Gly Glu Val Pro Gly Pro Arg Pro His Ser Gln Asn
305 310 315 320

Leu Pro Leu Asp Ser Asp Gln Glu Ser Gln Glu Ala Arg Leu Pro Glu
325 330 335

Arg Gly Thr Ala Leu Pro Thr Ala Arg Trp Pro Pro Arg Arg Ser Leu
340 345 350

Glu Arg Leu Pro Ser Pro Asp Pro Gly Ala Glu Gly His Gly Gln Ser
355 360 365

Arg Gln Ser Asp Gln Asp Ile Thr Lys Thr
370 375

<210> 21
<211> 356
<212> PRT
<213> Homo sapiens

<400> 21
Asp Arg Val Phe Gly Leu Glu Pro Pro Gly Thr Asn Met Ala Leu Val
1 5 10 15

Gly Leu Pro Gly Pro Asp Met Phe Cys Leu Phe His Gly Lys Arg Tyr
20 25 30

Ser Pro Gly Glu Ser Trp His Pro Tyr Leu Glu Pro Gln Gly Leu Met
35 40 45

Tyr Cys Leu Arg Cys Thr Cys Ser Glu Gly Ala His Val Ser Cys Tyr
50 55 60

Arg Leu His Cys Pro Pro Val His Cys Pro Gln Pro Val Thr Glu Pro
65 70 75 80

Gln Gln Cys Cys Pro Lys Cys Val Glu Pro His Thr Pro Ser Gly Leu
85 90 95

Arg Ala Pro Pro Lys Ser Cys Gln His Asn Gly Thr Met Tyr Gln His
100 105 110

Gly Glu Ile Phe Ser Ala His Glu Leu Phe Pro Ser Arg Leu Pro Asn
115 120 125

Gln Cys Val Leu Cys Ser Cys Thr Glu Gly Gln Ile Tyr Cys Gly Leu
130 135 140

Thr Thr Cys Pro Glu Pro Gly Cys Pro Ala Pro Leu Pro Leu Pro Asp
145 150 155 160

Ser Cys Cys Gln Ala Cys Lys Asp Glu Ala Ser Glu Gln Ser Asp Glu
165 170 175

Glu Asp Ser Val Gln Ser Leu His Gly Val Arg His Pro Gln Asp Pro
180 185 190

Cys Ser Ser Asp Ala Gly Arg Lys Arg Gly Pro Gly Thr Pro Ala Pro
195 200 205

Thr Gly Leu Ser Ala Pro Leu Ser Phe Ile Pro Arg His Phe Arg Pro
210 215 220

Lys Gly Ala Gly Ser Thr Thr Val Lys Ile Val Leu Lys Glu Lys His
225 230 235 240

Lys Lys Glu Asp Lys Ala Asp Pro Gly His Ser Glu Ile Ser Ser Thr
245 250 255

Arg Cys Pro Lys Ala Pro Gly Arg Val Leu Val His Thr Ser Val Ser
260 265 270

Pro Ser Pro Asp Asn Leu Arg Arg Phe Ala Leu Glu His Glu Ala Ser
275 280 285

Asp Leu Val Glu Ile Tyr Leu Trp Lys Leu Val Lys Gly Ile Phe His
290 295 300

Leu Thr Gln Ile Lys Lys Val Arg Lys Gln Asp Phe Gln Lys Glu Ala
305 310 315 320

Gln His Phe Arg Leu Leu Ala Gly Pro His Glu Gly His Trp Asn Val
325 330 335

Phe Leu Ala Gln Thr Leu Glu Leu Lys Val Thr Ala Ser Pro Asp Lys
340 345 350

Val Thr Lys Thr
355

<210> 22
<211> 397
<212> PRT
<213> Mouse

<400> 22
Phe Leu Tyr Ser Ser His Thr Ala Leu Pro Thr His Thr Ser Pro Lys
1 5 10 15

Val Xaa Glu Ser Pro Gly Gly Trp Leu Ala Lys Ser Leu Ser Val Xaa
20 25 30

Leu Leu Ile Ser Leu Arg Ile Ser Thr Ser Pro Thr Arg Phe Cys Val
35 40 45

Glu Pro Val Leu Ser Val Cys Leu Ser Val Cys Leu Ser Val Cys Leu
50 55 60

Ser Ala Cys Leu Ser Leu Ser Val Ser Val Cys Leu Cys Leu Ser Val
65 70 75 80

Cys Leu Cys Leu Ser Leu Ser Leu Cys Leu Ser Leu Cys Leu Cys Leu
85 90 95

Cys Leu Cys Leu Ser Leu Ser Leu Arg Ser Pro Leu Ala Phe Ser Ser
100 105 110

Arg Arg Leu Met Gln Pro Gly Trp Cys Ser Gln Leu Trp Pro Ile Pro
115 120 125

Gln Thr Ala Pro His Pro Ala Cys Cys Ser Gln Arg His Ser Gln Asp
130 135 140

Pro Cys Ser Glu Arg Arg Gly Pro Ser Thr Pro Ala Pro Thr Ser Leu
145 150 155 160

Ser Ser Pro Leu Gly Phe Ile Xaa Arg His Phe Gln Ser Val Gly Met
165 170 175

Gly Ser Thr Thr Ile Lys Ile Ile Leu Lys Glu Lys His Lys Lys Ala
180 185 190

Cys Thr His Asn Gly Lys Thr Tyr Ser His Gly Glu Val Trp His Pro
195 200 205

Thr Val Leu Ser Phe Gly Pro Met Pro Cys Ile Leu Cys Thr Cys Ile
210 215 220

Asp Gly Tyr Gln Asp Cys His Arg Val Thr Cys Pro Thr Gln Tyr Pro
225 230 235 240

Cys Ser Gln Pro Lys Lys Val Ala Gly Lys Cys Cys Lys Ile Cys Pro
245 250 255

Glu Asp Glu Ala Glu Asp Asp His Ser Glu Val Ile Ser Thr Arg Cys
260 265 270

Pro Lys Val Pro Gly Gln Phe Gln Val Tyr Thr Leu Ala Ser Pro Ser
275 280 285

Pro Asp Ser Leu His Arg Phe Val Leu Glu His Glu Ala Ser Asp Gln
290 295 300

Val Glu Met Tyr Ile Trp Lys Leu Val Lys Gly Ile Tyr His Leu Val
305 310 315 320

Gln Ile Lys Arg Val Arg Lys Gln Asp Phe Gln Lys Glu Val Gln Asn
325 330 335

Phe Arg Leu Leu Thr Gly Thr His Glu Gly Tyr Trp Thr Val Phe Leu
340 345 350

Ala Gln Ile Pro Glu Leu Lys Val Thr Ala Ser Pro Asp Lys Val Thr
355 360 365

Lys Thr Leu Gln Gly Pro Lys Glu Leu Gln Ile Arg Val Leu Leu Val
370 375 380

Leu Leu Leu Tyr Ile Asn Lys Glu Val Ala Leu Pro Phe
385 390 395

INTERNATIONAL SEARCH REPORT

[redacted] onal Application No
PCT/IL 00/00736

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/12 C07K14/475 C07K16/22 C12Q1/68 G01N33/68
G01N33/53 A61K38/17

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, EMBL, CHEM ABS Data, BIOSIS, WPI Data, SCISEARCH, EMBASE, BIOTECHNOLOGY ABS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 98 40483 A (HUMAN GENOME SCIENCES INC ;GREENE JOHN M (US); LI YI (US); ROSEN C) 17 September 1998 (1998-09-17) page 148 -page 149; claim 1 page 164 -page 165 ---	1-13,15
X	WO 99 54353 A (SCHMITT ARMIN ;SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN BERND) 28 October 1999 (1999-10-28) abstract; figures SEQ.ID.19,113 ---	1-13,15

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
22 March 2001	05/04/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Gurdjian, D

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL 'Online! 23 June 1999 (1999-06-23) ROBERT STRAUSBERG: "similar to to SW:CA11_CHICK P02457 PROCOLLAGEN ALPHA 1(I) CHAIN PRECURSOR." retrieved from EBI Database accession no. AI739159 XP002163678 abstract --- DATABASE EMBL 'Online! 20 August 1999 (1999-08-20) DUESTERHOEFT A. ET AL.: "Homo sapiens mRNA; cDNA DKFZp586N2124 (from clone DKFZp586N2124)" retrieved from EBI Database accession no. AL110168 XP002163679 abstract --- DATABASE SWALL 'Online! 1 November 1999 (1999-11-01) PEARCE A.: "DA141H5.1 (C-TERMINAL PART OF A CHORDIN LIKE PROTEIN WITH VON WILLEBRAND FACTOR TYPE C DOMAINS) " retrieved from EBI Database accession no. Q9Y3H7 XP002163680 abstract	1-12
X		1-12
A		1-12
P,X	WO 00 09551 A (GENETICS INST) 24 February 2000 (2000-02-24) page 99 -page 100; claim 32	1,6,8-17
P,X	WO 00 12708 A (BAKER KEVIN ;GENENTECH INC (US); GODDARD AUDREY (US); GURNEY AUSTI) 9 March 2000 (2000-03-09) claim 12; figures SEQ.ID.141,142	1-13,15
P,X	WO 99 57132 A (GENETICS INST) 11 November 1999 (1999-11-11) page 430 -page 432; claim 85; figure SEQ.ID.76	1-13,15
E	WO 00 70049 A (INCYTE GENOMICS INC ;PATTERSON CHANDRA (US); AZIMZAI YALDA (US); Y) 23 November 2000 (2000-11-23) figure SEQ.ID.39	1-13,15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/IL 00/00736

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9840483 A	17-09-1998	AU 6552198 A	29-09-1998
		EP 0973892 A	26-01-2000
		EP 1039801 A	04-10-2000
		WO 9854963 A	10-12-1998
		AU 2306499 A	05-07-1999
		EP 1040117 A	04-10-2000
		WO 9931117 A	24-06-1999
WO 9954353 A	28-10-1999	DE 19817946 A	21-10-1999
		EP 1071777 A	31-01-2001
WO 0009551 A	24-02-2000	AU 4071199 A	23-11-1999
		AU 5475199 A	06-03-2000
		EP 1077991 A	28-02-2001
		WO 9957132 A	11-11-1999
WO 0012708 A	09-03-2000	AU 5590899 A	21-03-2000
		AU 6041399 A	10-04-2000
		WO 0017353 A	30-03-2000
WO 9957132 A	11-11-1999	AU 4071199 A	23-11-1999
		EP 1077991 A	28-02-2001
		AU 5475199 A	06-03-2000
		WO 0009551 A	24-02-2000
WO 0070049 A	23-11-2000	AU 5151100 A	05-12-2000

PATENT COOPERATION TREATY

PCT

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

(PCT Article 36 and Rule 70)

8

Applicant's or agent's file reference 129275.4 DAB	**FOR FURTHER ACTION**	See Notification of Transmittal of International Preliminary Examination Report (Form PCT/IPEA/416)
International application No. PCT/IL00/00736	International filing date *(day/month/year)* 10/11/2000	Priority date *(day/month/year)* 10/11/1999
International Patent Classification (IPC) or national classification and IPC C12N15/12		
Applicant COMPUGEN LTD. et al		

1. This international preliminary examination report has been prepared by this International Preliminary Examining Authority and is transmitted to the applicant according to Article 36.

2. This REPORT consists of a total of 8 sheets, including this cover sheet.

   ☒ This report is also accompanied by ANNEXES, i.e. sheets of the description, claims and/or drawings which have been amended and are the basis for this report and/or sheets containing rectifications made before this Authority (see Rule 70.16 and Section 607 of the Administrative Instructions under the PCT).

   These annexes consist of a total of 4 sheets.

3. This report contains indications relating to the following items:

   - I ☒ Basis of the report
   - II ☐ Priority
   - III ☒ Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
   - IV ☒ Lack of unity of invention
   - V ☒ Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations suporting such statement
   - VI ☐ Certain documents cited
   - VII ☐ Certain defects in the international application
   - VIII ☐ Certain observations on the international application

Date of submission of the demand	Date of completion of this report
21/05/2001	11.03.2002
Name and mailing address of the international preliminary examining authority: European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Tx: 523656 epmu d Fax: +49 89 2399 - 4465	Authorized officer Nichogiannopoulou, A Telephone No. +49 89 2399 8054

INTERNATIONAL PRELIMINARY EXAMINATION REPORT

International application No. PCT/IL00/00736

I. Basis of the report

1. With regard to the **elements** of the international application *(Replacement sheets which have been furnished to the receiving Office in response to an invitation under Article 14 are referred to in this report as "originally filed" and are not annexed to this report since they do not contain amendments (Rules 70.16 and 70.17))*:
Description, pages:

1-53,55	as originally filed				
54	as received on	26/02/2002	with letter of	04/02/2002	

Claims, No.:

1-7	as received on	26/02/2002	with letter of	04/02/2002	

Drawings, sheets:

1/116-116/116	as originally filed

Sequence listing part of the description, pages:

1-24, as originally filed

2. With regard to the **language**, all the elements marked above were available or furnished to this Authority in the language in which the international application was filed, unless otherwise indicated under this item.

   These elements were available or furnished to this Authority in the following language: , which is:

   ☐ the language of a translation furnished for the purposes of the international search (under Rule 23.1(b)).

   ☐ the language of publication of the international application (under Rule 48.3(b)).

   ☐ the language of a translation furnished for the purposes of international preliminary examination (under Rule 55.2 and/or 55.3).

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, the international preliminary examination was carried out on the basis of the sequence listing:

   ☒ contained in the international application in written form.

   ☒ filed together with the international application in computer readable form.

   ☐ furnished subsequently to this Authority in written form.

   ☐ furnished subsequently to this Authority in computer readable form.

   ☐ The statement that the subsequently furnished written sequence listing does not go beyond the disclosure in the international application as filed has been furnished.

   ☐ The statement that the information recorded in computer readable form is identical to the written sequence

listing has been furnished.

4. The amendments have resulted in the cancellation of:

   ☐ the description, pages:

   ☐ the claims, Nos.:

   ☐ the drawings, sheets:

5. ☐ This report has been established as if (some of) the amendments had not been made, since they have been considered to go beyond the disclosure as filed (Rule 70.2(c)):

   (Any replacement sheet containing such amendments must be referred to under item 1 and annexed to this report.)

6. Additional observations, if necessary:

III. Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

1. The questions whether the claimed invention appears to be novel, to involve an inventive step (to be non-obvious), or to be industrially applicable have not been examined in respect of:

   ☐ the entire international application.

   ☒ claims Nos. 2-7.

because:

   ☒ the said international application, or the said claims Nos. 2-7 relate to the following subject matter which does not require an international preliminary examination (*specify*):
   see separate sheet

   ☐ the description, claims or drawings (*indicate particular elements below*) or said claims Nos. are so unclear that no meaningful opinion could be formed (*specify*):

   ☐ the claims, or said claims Nos. are so inadequately supported by the description that no meaningful opinion could be formed.

   ☐ no international search report has been established for the said claims Nos. .

2. A meaningful international preliminary examination cannot be carried out due to the failure of the nucleotide and/or amino acid sequence listing to comply with the standard provided for in Annex C of the Administrative Instructions:

   ☐ the written form has not been furnished or does not comply with the standard.

   ☐ the computer readable form has not been furnished or does not comply with the standard.

IV. Lack of unity of invention

1. In response to the invitation to restrict or pay additional fees the applicant has:

   ☐ restricted the claims.

   ☐ paid additional fees.

   ☐ paid additional fees under protest.

   ☐ neither restricted nor paid additional fees.

2. ☒ This Authority found that the requirement of unity of invention is not complied and chose, according to Rule 68.1, not to invite the applicant to restrict or pay additional fees.

3. This Authority considers that the requirement of unity of invention in accordance with Rules 13.1, 13.2 and 13.3 is

   ☐ complied with.

   ☒ not complied with for the following reasons:
   see separate sheet

4. Consequently, the following parts of the international application were the subject of international preliminary examination in establishing this report:

   ☒ all parts.

   ☐ the parts relating to claims Nos. .

V. Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	1-7
	No:	Claims	
Inventive step (IS)	Yes:	Claims	1-7
	No:	Claims	
Industrial applicability (IA)	Yes:	Claims	1
	No:	Claims	

2. Citations and explanations
   see separate sheet

Re Item I

Basis of the opinion

1. The amendments filed with the letter of 04.02.2002 are formally allowable under Article 34(2)(b) PCT because they do not introduce subject-matter extending beyond the content of the application as filed. It is herewith noted however, that failure of the applicant to submit a detailed basis for the filed amendments rendered the task of identifying such a basis particularly tedious.

Re Item III

Non-establishment of opinion with regard to novelty, inventive step and industrial applicability

1. New claim 7 appears to be dependent on new claim 1, which would make little technical sense and subsequent examination impossible. The IPEA has examined the claim upon the assumption that the correct dependancy is on new claim 6.

2. Claims 2-7 -since they concern *in vivo* methods- relate to subject-matter considered by this Authority to be covered by the provisions of Rule 67.1(iv) PCT. Consequently, no opinion will be formulated with respect to the industrial applicability of the subject-matter of these claims (Article 34(4)(a)(i) PCT).

Re Item IV

Lack of unity of invention

1. The present application relates to 11 different nucleic acid sequences (SEQ ID Nos:1-11) and the proteins they putatively encode (SEQ ID Nos:12-24). The common feature among these sequences, is that they have homology to the known chordin. Since no further technical feature could be attributed to the claimed sequences, this common concept is found to lack an inventive step. Accordingly it cannot serve as the unifying feature linking the 11 claimed sequences. The IPEA therefore is of the opinion that the present application lacks unity pursuant to Rule 13 PCT, and considers that it relates to eleven distinct inventions. This objection is currently not being pursued. Should the application enter the European phase, an objection under the corresponding EPC article will be raised.

Re Item V

Reasoned statement under Article 35(2) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Priority

1.1. The present application validly claims priority from 10.11.1999 in as far as claims to SEQ ID Nos:1 and 12 are covered and from 28.12.1999 for claims to SEQ ID Nos:1-4 and 12-15. Any documents cited in the International Search Report as P documents have therefore not been considered as comprised in the prior art relevant for said sequences.
Priority has however been invalidly claimed for SEQ ID Nos:5-11 (DNA) and 16-22 (Protein) for which the present application represents the first disclosure.

1.2. The following documents were published prior to the international filing date but later than the priority date claimed (P-documents):

P1: WO 00 09551 A (GENETICS INST) 24 February 2000 (2000-02-24)
P2: WO 00 12708 A (BAKER KEVIN ;GENENTECH INC (US); GODDARD AUDREY (US); GURNEY AUSTI) 9 March 2000 (2000-03-09)
P3: WO 99 57132 A (GENETICS INST) 11 November 1999 (1999-11-11)

The content of these documents is considered as comprised in the state of the art for the purposes of examination of claims pertaining to SEQ ID Nos:5-11 and 16-22.

2. Conflicting European Application

Document WO 00 70049 has been published after the filing date of the present application (publication date: 23.11.2000, filing date: 19.05.2000, priority date: 19.05.1999) and does therefore not constitute prior art in the meaning of Rule 64(1)(b) PCT. Said document will, however become of relevance for novelty assessment should the application enter the European phase.

3. Reference is made to the following documents:

D1: WO 98 40483 A (HUMAN GENOME SCIENCES INC) 17 September 1998 (1998-09-17)
D2: WO 99 54353 A (SCHMITT A et al) 28 October 1999 (1999-10-28)
D3: DATABASE EMBL [Online] 23 June 1999 (1999-06-23) ROBERT STRAUSBERG: 'similar to SW:CA11_CHICK P02457 PROCOLLAGEN ALPHA 1(I) CHAIN PRECURSOR.' retrieved from EBI Database accession no. AI739159 XP002163678

4. **Novelty** (Article 33(2) PCT)

The present application discloses 11 DNA sequences and the proteins they putatively encode.

4.1. **D1** discloses genes encoding secreted proteins. Two of these have 99.2% and 99% identity with SEQ ID Nos:2 and 3 of the present application over 1703 and 1144

nucleotides respectively. For the sake of completion it is noted that **D2** also discloses nucleic acids and proteins with >95% identity to the nucleic acids of SEQ ID Nos:2, 4, 5, 6, 8, 10 and >90% identity with the proteins of SEQ ID Nos:13-19 of the present application. **D3** is a database entry with 100% identity with SEQ ID No:9 of the present application over 482 nucleotides and >99% identity with the nucleotides of SEQ ID Nos:1-8.

4.2. The new set of claims discloses the therapeutic and diagnostic utility of the disclosed sequences. Such utility has been neither disclosed nor suggested in the available prior art, so that new claims 1-7 would appear to satisfy the novelty and inventive step requirements set out in Articles 33(2) and (3) PCT.

5. **Industrial applicability** (Article 33(4) PCT)
The subject-matter of the new claims for which an opinion has been established (see item III) appears to be industrially applicable under the terms of Article 33(4) PCT.

variant 1 molecule is presented in figure 23.

As shown in Fig 23, lanes A3, B3, C3, COS7 untransfected cells (referred to as *Mock)* do not express CLH endogenously. CLH was over expressed only in the cells transfected with pCDNA3 carrying CLH gene Fig 23 , lane C1 and not in the cells transfected with pCDNA3 Fig 23 , lane C2. Moreover, high levels of secreted protein were detected in the medium of CLH transfected cells following 48hr and 72 hr (Fig 23 Lanes A1 and lane B1 respectively), and not in the cells transfected with pCDNA3 (Fig 23 Lanes A2 and lane B2 respectively).

EXAMPLE IX: Immunohistochemical localization of CLH protein in different human tissues:

Immunohistochemical staining was performed on different human micron sections using the anti-LM antibodies (Fig 24 right column letters with prime) indicated compared to the pre-immune rabbit's serum (Fig 24, left columns, indicated in normal letters). CLH was found to be expressed in different epithelial tissues (Fig.24 a', b', c', d', e', f', g') and localized mainly in the secreting cells.

Expression of CLH was detected in ductal epithelium of the breast. Breast carcinoma was positively stained both in the regions of ductal carcinoma (Fig. 24 a') in situ (DCIS) and of invasive ductal carcinoma (Fig. 24b). Secreting cells in benign prostatic hyperplasia (BPH) and prostate carcinoma sections were also positively stained Fig. 24, c', d', respectively.

CLH was localized to the transitional epithelium in the bladder (Fig 24 e'). The internal female genitalia (fallopian tube, endocervical glands and the uterus) which evolved from the same embryonic precursor - the mullerian duct, showed positive staining (Fig.24, e'). Expression of CLH was localized in the lining epithelium of the fallopian tube (Fig 24, f'), in the endocervical glands (Fig 24, g') and in the normal and endometrial carcinoma of the uterus (Fig. 24, h' and i', respectively). However, in the region of the mucinous metaplasia in the endometrial carcinoma, negative staining of CLH was observed

CLAIMS:

1. Use of an active ingredient selected from:

- expression vector comprising a nucleic acid sequence of SEQ ID NO:1 to SEQ ID NO:11 or a nucleic acid sequence complementary thereto, and a control element for the expression of the nucleic acid sequence in a host cell,
- an amino acid sequence coded by the amino acid sequence of SEQ ID NO:1 to SEQ ID NO:11 or an amino acid sequence of any one of SEQ ID NO:12 to SEQ ID NO:22, and
- a purified antibody which binds specifically to said amino acid sequence,

for the preparation of a pharmaceutical composition for the treatment of a disease or disorder being one or more of diseases manifested in non-normal bone formation and non-normal bone modeling; bone injuries; neuronal diseases of the CNS; neurodegenerative diseases; and diseases involving non-normal development of neurons.

2. A method of treatment of a disease or disorder being one or more of diseases manifested in non-normal bone formation and non-normal bone modeling; bone injuries; neuronal diseases of the CNS; neurodegenerative diseases; and diseases involving non-normal development of neurons; the method comprising administering to a needy subject an effective amount of an active agent selected from:

- expression vector comprising a nucleic acid sequence of SEQ ID NO:1 to SEQ ID NO:11 or a nucleic acid sequence complementary thereto, and a control element for the expression of the nucleic acid sequence in a host cell,
- an amino acid sequence coded by the amino acid sequence of SEQ ID NO:1 to SEQ ID NO:11 or an amino acid sequence of any one of SEQ ID NO:12 to SEQ ID NO:22, and

- a purified antibody which binds specifically to said amino acid sequence.

3. A method for diagnosing a disease or disorder being one or more of diseases manifested in non-normal bone formation and non-normal bone modeling; bone injuries; neuronal diseases of the CNS; neurodegenerative diseases; and diseases involving non-normal development of neurons in a subject; the method comprising:

(i) obtaining a biological sample from the subject;

(ii) hybridizing to a nucleic acid material of said biological sample a nucleic acid sequence of SEQ ID NO:1 to SEQ ID NO:11 or a nucleic acid sequence complementary thereto; and

(iii) detecting nucleic acid hybridization complex.

4. A method according to Claim 3, wherein the nucleic acid material of said biological sample includes mRNA transcripts.

5. A method according to Claim 3, wherein the nucleic acid sequence is immobilized on a nucleic acid chip.

6. A method for diagnosing a disease or disorder being one or more of diseases manifested in non-normal bone formation and non-normal bone modeling; bone injuries; neuronal diseases of the CNS; neurodegenerative diseases; and diseases involving non-normal development of neurons in a subject; the method comprising:

(i) obtaining a biological sample from the subject;

(ii) detecting the presence of a polypeptide having the amino acid sequence of any one of SEQ ID NO:12 to SEQ ID NO:22 in the sample.

7. A method according to Claim 1, comprising:

(i) obtaining a biological sample from the subject;

(ii) contacting said sample with purified antibodies that bind specifically to the sequences of any one of SEQ ID NO:12 to SEQ ID NO:22

(iii) detecting complexes between said antibodies and antigens in the sample.